



# Large Independent Sets from Local Considerations

Matija Bucić<sup>1,2</sup> · Benny Sudakov<sup>3</sup>

Received: 12 May 2021 / Revised: 17 October 2022 / Accepted: 11 January 2023 /  
Published online: 4 May 2023  
© The Author(s) 2023

## Abstract

The following natural problem was raised independently by Erdős–Hajnal and Linial–Rabinovich in the early '90s. How large must the independence number  $\alpha(G)$  of a graph  $G$  be whose every  $m$  vertices contain an independent set of size  $r$ ? In this paper, we discuss new methods to attack this problem. The first new approach, based on bounding Ramsey numbers of certain graphs, allows us to improve the previously best lower bounds due to Linial–Rabinovich, Erdős–Hajnal and Alon–Sudakov. As an example, we prove that any  $n$ -vertex graph  $G$  having an independent set of size 3 among every 7 vertices has  $\alpha(G) \geq \Omega(n^{5/12})$ . This confirms a conjecture of Erdős and Hajnal that  $\alpha(G)$  should be at least  $n^{1/3+\varepsilon}$  and brings the exponent halfway to the best possible value of  $1/2$ . Our second approach deals with upper bounds. It relies on a reduction of the original question to the following natural extremal problem. What is the minimum possible value of the 2-density (The 2-density of a graph  $H$  is defined as  $m_2(H) := \max_{H' \subseteq H, |H'| \geq 3} \frac{e(H')-1}{|H'|-2}$ ) of a graph on  $m$  vertices having no independent set of size  $r$ ? This allows us to improve previous upper bounds due to Linial–Rabinovich, Krivelevich and Kostochka–Jancey. As part of our arguments, we link the problem of Erdős–Hajnal and Linial–Rabinovich and our new extremal 2-density problem to a number of other well-studied questions. This leads to many interesting directions for future research.

**Keywords** Local to global phenomenon · Ramsey Theory · Erdos-Rogers problem · 2-density · Lovasz Local Lemma

---

The author Matija Bucić is supported in part by the NSF Grant CCF-1900460. The author Benny Sudakov is supported in part by SNSF Grant 200021\_196965.

---

✉ Matija Bucić  
matija.bucic@ias.edu

Benny Sudakov  
benjamin.sudakov@math.ethz.ch

<sup>1</sup> School of Mathematics, Institute for Advanced Study, Princeton, USA

<sup>2</sup> Department of Mathematics, Princeton University, Princeton, USA

<sup>3</sup> Department of Mathematics, ETH, Zürich, Switzerland

**Mathematics Subject Classification** 05D10 · 05C55 · 05C35 · 05C15 · 05C65 · 05D40

## 1 Introduction

In this paper, we study the following classical problem. If we know that any  $m$  vertices of a graph contain an independent set of order  $r$  how large can the independence number of the whole graph be? The study of this problem for a specific choice of parameters dates back almost 60 years, with the first published result being due to Erdős and Rogers [13] in 1962.

Over the years this problem attracted a lot of attention. Originally the focus was on the instance of the problem in which we keep the sizes of independent sets we want to find locally and in the whole graph to be fixed and small. In other words, if we forbid in  $G$  an independent set of size  $s$ , how big a subset of vertices one can find without an independent set of size  $r$ ? Choosing  $r = 2$  precisely recovers the usual Ramsey problem and was in fact the original motivation behind the general question. This question became known as the Erdős–Rogers problem and has been extensively studied, for some examples see [9, 11, 13, 20, 25, 37–39] and a recent survey [10] due to Dudek and Rödl.

In the early 90’s Erdős and Hajnal [12] and independently Linial and Rabinovich [30] propose changing the perspective and fixing the local parameters  $m$  and  $r$  instead. In other words, asking what can be said about the independence number of the whole graph if we know that any small number of vertices  $m$  contain an independent set of size  $r$ . This frames the problem squarely under the so-called local–global principle, stating that one can obtain global understanding of a structure from having a good understanding of its local properties, or vice versa. This phenomenon has been ubiquitous in many areas of mathematics and beyond, see e.g. [5, 18, 19, 29]. In fact, one can define an  $m$ -local independence number  $\alpha_m(G)$  of a graph  $G$  to be the minimum independence number we can find among subgraphs of  $G$  on  $m$  vertices and the problem becomes relating the local independence number to the independence number of  $G$  itself, the “global” independence number. In particular, we are interested in the smallest possible size of  $\alpha(G)$  in an  $n$ -vertex graph satisfying  $\alpha_m(G) \geq r$ .

In this paper we discuss two new approaches for attacking this problem, which allow us to significantly improve previously best-known bounds due to Linial and Rabinovich [30], Erdős–Hajnal [12], Alon and Sudakov [4], Krivelevich [26] and Kostochka and Jancey [24]. In the case of lower bounds, we improve their results for at least half of the possible choices of  $m$  and  $r$  and in the case of upper bounds for essentially all choices. Moreover, we believe that both approaches have the potential for further improvements.

The initial approach of Linial and Rabinovich [30] and independently Alon and Sudakov [4] reduces the lower bound problem to the question of bounding from above Ramsey numbers of a clique of size  $k = \lceil \frac{m}{r-1} \rceil$  vs a large independent set. Our new idea is that one can find other “forbidden” graphs whose Ramsey numbers perform better. For this to work we need to obtain upper bounds on the Ramsey numbers of our new graphs vs a large independent set, which often turns out to be an interesting

problem in its own right. See the beginning of Sect. 2.1 for a more detailed illustration of our new approach.

The above-introduced parameter  $k$  controls in large part the known lower bounds for  $\alpha(G)$  among all graphs satisfying  $\alpha_m(G) \geq r$ . Linial and Rabinovich [30] determine the answer precisely if  $k \leq 2$ , i.e. for  $m \leq 2r - 2$ . For  $k = 3$ , they show that an  $n$  vertex graph satisfying  $\alpha_m(G) \geq r$  must have  $\alpha(G) \geq n^{1 - \frac{2}{r-1} - o(1)}$  if  $m = 2r - 1$  and  $\alpha(G) \geq \Omega(n^{1/2})$  for the rest of the range  $m \leq 3r - 3$ . Our first result improves the exponent in their bounds for the first half of this range. Moreover, the improvement in the exponent is by a constant factor independent of  $r$ , unless  $m = 2r - 1$ .

**Proposition 1.1** *Let  $m = 2r - 2 + t$  for  $1 \leq t \leq r - 1$ . Then any  $n$ -vertex graph  $G$  satisfying  $\alpha_m(G) \geq r$  has  $\alpha(G) \geq \Omega(n^{1-1/\ell})$ , where  $\ell = \lfloor \frac{r-1}{t} \rfloor + 1$ .*

In the general case of  $k \geq 4$ , Linial and Rabinovich and independently Alon and Sudakov show that an  $n$ -vertex graph satisfying  $\alpha_m(G) \geq r$  must have  $\alpha(G) \geq \Omega(n^{\frac{1}{k-1}})$ . We improve the exponent in these bounds for the first half of the range for any  $k$ .

**Theorem 1.2** *Let  $k = \lceil \frac{m}{r-1} \rceil$  and let us assume  $m \leq (k - \frac{1}{2})(r - 1)$ . Then any  $n$ -vertex graph  $G$  satisfying  $\alpha_m(G) \geq r$  has  $\alpha(G) \geq \Omega(n^{\frac{1}{k-3/2}})$ .*

Going beyond  $k - 2$  in the denominator of the exponent in the above theorem seems likely to require an improvement over the best-known upper bounds on Ramsey numbers, which have not seen an improvement in the exponent since the initial paper of Erdős and Szekeres [14] from 1935. This means our result is in some sense halfway between the previously best bound and the Ramsey barrier.

The key part of the above result is actually the special case of  $r = 3$ . This is due to an easy observation which allows us to generalise any improvement in this case to the first half of the range as above, for any  $r$ . The first interesting instance here, which actually lead us to the general improvements above, is  $m = 7$  and  $r = 3$  in which case we can obtain an even better bound. Studying this case was explicitly proposed by Erdős and Hajnal [12] who observed that any graph  $G$  on  $n$  vertices with  $\alpha_7(G) \geq 3$  must have  $\alpha(G) \geq \Omega(n^{1/3})$  and that such a graph  $G$  exists with  $\alpha(G) \leq O(n^{1/2})$ . They conjectured that neither of these bounds is tight. Our next result confirms their first conjecture.

**Theorem 1.3** *Any  $n$ -vertex graph  $G$  with  $\alpha_7(G) \geq 3$  has  $\alpha(G) \geq n^{5/12 - o(1)}$ .*

By the aforementioned observation, this actually gives the same improved bound for the first half of the range for any instance with  $k = 4$ , i.e. for  $3r - 2 \leq m \leq 3.5(r - 1)$ .

To prove an upper bound on the minimum possible  $\alpha(G)$  among all graphs with  $\alpha_m(G) \geq r$ , one needs to find a graph which has small independence number while having big independent sets spread around everywhere. Given the close relation of our problem to Ramsey numbers, random graphs are natural candidates for such examples. A binomial random graph  $\mathcal{G}(n, p)$  is an  $n$ -vertex graph in which we include every possible edge independently with probability  $p$ . Understanding  $\alpha_m(G)$  in  $G \sim \mathcal{G}(n, p)$  turns out to be an interesting problem in its own right. Observe that the requirement  $\alpha_m(G) \geq r$  may be rephrased as stating that  $G$  contains no copy of an  $m$ -vertex graph

$H$  with  $\alpha(H) \leq r - 1$  as a subgraph. A standard application of Lovász local lemma tells us that if we are only forbidding a single graph  $H$ , then the largest  $p$  we can take is controlled by the 2-density of  $H$  (see Sect. 3 for the definition of 2-density and more details). If we are instead forbidding a family of graphs the correct parameter turns out to be the minimum of the 2-densities over all graphs in our family. This reduces our problem to the following natural extremal question, which we propose to study. What is the minimum value of the 2-density of an  $m$ -vertex graph  $H$  with  $\alpha(H) \leq r - 1$ ? If we denote the answer to this question by  $M(m, r)$  the above discussion leads us to the following reduction.

**Proposition 1.4** *Let  $m, r$  be fixed,  $m \geq 2r - 1 \geq 3$  and  $M = M(m, r)$ . Then for any  $n$  there exists an  $n$ -vertex graph  $G$  with  $\alpha_m(G) \geq r$  and  $\alpha(G) \leq n^{1/M+o(1)}$ .*

The value of  $M(m, r)$ , and hence also our upper bounds for the local to global independence number problem, are mostly controlled by the same parameter  $k = \lceil \frac{m}{r-1} \rceil$  as before. Some intuition behind this, suggested by Linial and Rabinovich [30], is that a natural example of an  $m$  vertex graph with independence number at most  $r - 1$  is a vertex disjoint union of  $r - 1$  cliques with sizes as equal as possible (in other words complement of a Turán graph on  $m$  vertices with no clique of size  $r$ ). This graph clearly has no independent set of size  $r$  and we picked the clique sizes as equal as possible in order to minimise the 2-density. Our parameter  $k$  is simply the size of a largest clique in this example.

Turning to the results, we start once again with the range  $2r - 1 \leq m \leq 3r - 3$ , i.e.  $k = 3$ . Here Linial and Rabinovich show that there exist  $n$ -vertex graphs  $G$  satisfying  $\alpha_m(G) \geq r$  and  $\alpha(G) \leq n^{1-1/(8r-4)}$ . We improve the exponent in this bound for the whole range. Moreover, the improvement in the exponent is by a constant factor independent of  $r$ , towards the end of the range.

**Proposition 1.5** *Let  $m \geq 2r - 1 \geq 3$ , for any  $n$  there exists an  $n$ -vertex graph  $G$  satisfying  $\alpha_m(G) \geq r$  with  $\alpha(G) \leq n^{1-\frac{1}{2r-2}+o(1)}$  and if  $m \geq 3r - 4 > 2$  with  $\alpha(G) \leq n^{\frac{3}{5}+\frac{2}{5r-13}+o(1)}$ .*

These bounds follow from our results on  $M(2r - 1, r)$ , which we determine precisely and  $M(3r - 4, r)$  which we determine up to lower order terms. This means that in terms of using random graphs as examples, these bounds are essentially best possible for  $m = 2r - 1, 3r - 4$ . We can obtain a constant factor improvement in the exponent for about 1/3 of the range, but since we believe our current argument does not give the best possible answer, in terms of  $M(m, r)$ , for the whole range we leave this open for future research.

The problem of determining  $M(3r - 4, r)$  is closely related to a well-studied problem of finding large independent sets in sparse triangle-free graphs. Perhaps the most famous result in this direction is due to Ajtai–Komlós–Szemerédi [1] and Shearer [34], but for our problem earlier results of Staton [35] and Jones [21] turn out to be more relevant. These results are part of a very active research area of studying graphs having no cliques of size  $k$  nor independent sets of size  $r$  but which have potentially much fewer vertices than the corresponding Ramsey number  $R(k, r)$ . Our problem of lower bounding  $M(m, r)$  falls under this framework since we can always assume that

our graphs, in addition to having no independent sets of size  $r$ , are also  $K_k$ -free or the 2-density is already large. We point the interested reader to classical papers [2, 35] and numerous papers citing them.

We now turn to the general case of  $k \geq 4$ . Let us begin with the initial instance, so when  $r = 3$ . Unfortunately, here the results for  $r = 3$  do not immediately generalise as they did in the case of lower bounds. They do however provide a starting point, which serves as a basis for more general results. Here we determine  $M(m, 3)$  precisely for all  $m$ , which allows us to improve exponents in the previously best bounds of Linial and Rabinovich [30]. They showed there are  $n$ -vertex graphs  $G$  with  $\alpha_m(G) \geq 3$  and  $\alpha(G) \leq n^{\frac{4+o(1)}{m-4/(m-2)}}$  if  $m$  is even, and  $\alpha(G) \leq n^{\frac{4+o(1)}{m-3/(m-2)}}$  if  $m$  is odd.

**Theorem 1.6** *For any  $n$  there exists an  $n$ -vertex graph  $G$  satisfying  $\alpha_m(G) \geq 3$  with  $\alpha(G) \leq n^{\frac{4+o(1)}{m+2}}$  if  $m$  is even, and  $\alpha(G) \leq n^{\frac{4+o(1)}{m+3-13/\sqrt{m}}}$  when  $m \geq 5$  is odd (here both terms  $o(1) \rightarrow 0$  as  $n \rightarrow \infty$ ).*

We remark that in the even case any improvement of our exponent, in terms of  $m$ , provably leads to improvement over the best-known lower bounds on Ramsey numbers and in the odd case without improving the Ramsey numbers one can only improve the term  $13/\sqrt{m}$ .

Once again our arguments in this particular regime show that the problem of determining  $M(m, r)$  is related to yet another well-studied problem. Namely, the stability problem for Turán’s theorem first considered by Erdős, Győri and Simonovits in [16]. While one can use their results to obtain good bounds on  $M(m, 3)$  obtaining precise answers requires a different, more careful argument.

In the fully general case we determine  $M(m, r)$  up to lower order terms (where  $r$  is considered fixed and  $m$  large), giving us the following result.

**Theorem 1.7** *Let us assume  $m$  is sufficiently larger than  $r$  and set  $k = \lceil \frac{m}{r-1} \rceil$ . Then there exists  $c_r > 0$  such that for any  $n$  there exists an  $n$ -vertex graph  $G$  satisfying  $\alpha_m(G) \geq r$  with  $\alpha(G) \leq n^{\frac{2+o(1)}{k+1-c_r/\sqrt{k}}}$ .*

This improves previous upper bounds of Linial and Rabinovich from roughly  $n^{\frac{2}{k-1}}$  to roughly  $n^{\frac{2}{k+1}}$  and is once again essentially best possible assuming lower bounds on Ramsey numbers are tight.

While the above result requires  $m$  to be large compared to  $r$  some of our ideas apply for any choice of the parameters. To illustrate this we consider the case  $m = 20, r = 5$  which was used by various researchers as a benchmark to compare their methods. Here Linial and Rabinovich show there are graphs  $G$  having  $\alpha_{20}(G) \geq 5$  and  $\alpha(G) \leq n^{18/39+o(1)}$ . Krivelevich [26] improved this to  $\alpha(G) \leq n^{14/33+o(1)}$ . He obtains this as an application of his result on the minimum number of edges in colour-critical graphs. The best possible bound using this approach was later obtained by Kostochka and Jancey [24] who showed  $\alpha(G) \leq n^{18/43+o(1)}$ . For comparison  $39/18 \approx 2.17, 33/14 \approx 2.36$  and  $43/18 \approx 2.39$ , while our methods allow us to improve this to 3. That is, there exists a graph  $G$  with  $\alpha_{20}(G) \geq 5$  which has  $\alpha(G) \leq n^{1/3+o(1)}$  and once again this is best possible (up to the  $o(1)$  term) without improving the lower bounds on Ramsey numbers  $R(5, s)$ .

In addition to the above applications and connections, another reason which makes the study of  $M(m, r)$  interesting is its relation to a random graph process. For a graph property  $\mathcal{P}$  the random graph process with respect to  $\mathcal{P}$  starts with an empty graph and iteratively adds a new uniformly random edge for as long as this does not violate  $\mathcal{P}$ . Random graph processes have been extensively studied for a variety of properties and have found numerous applications (see e.g. [7, 8, 17, 22, 27, 32] and references therein). In our setting  $M(m, r)$  controls the final density of the random process with respect to the  $m$ -local  $r$ -independence property  $\alpha_m \geq r$ . So  $M(m, r)$  essentially controls the behaviour of this random process.

**Organisation.** We will prove our lower bound results in Sect. 2. We start this section by proving Proposition 1.1. We continue by formalising in the form of Lemma 2.1 the above-mentioned reduction of our general lower bound result, Theorem 1.2, to the case  $r = 3$ . We then focus on this case in Sect. 2.1 to complete the proof of Theorem 1.2. In Sect. 2.2 we prove our stronger bound in the  $(m, r) = (7, 3)$  case raised by Erdős and Hajanal, namely Theorem 1.3. In Sect. 3 we prove our upper bound results. We begin by proving our reduction to the 2-density Turán problem, namely Proposition 1.4. We then switch the focus to proving our results concerning this 2-density Turán problem in Sect. 3.1. We begin by proving our results in the triangle-free regime in Sect. 3.1.1 providing us with a proof of Proposition 1.5. In Sect. 3.1.2 to solve the independence number two case which provides us with a proof of Theorem 1.6. We complete the section with our asymptotic solution to the general problem which gives a proof of Theorem 1.7. In Sect. 4 we give some concluding remarks and open problems as well as the summary of our results. We also include three appendices. In Appendix A we prove a slight modification of a result of [21] which we need in the proof of the second part of Proposition 1.5. In Appendix B we completely solve a benchmark case  $(m, r) = (20, 5)$  of the Turán 2-density problem. Finally, in Appendix A we prove some upper bounds for the Turán 2-density problem which establish the tightness of our results in this direction but are not necessary for our reduction to the local to global independence number problem.

**Notation.** We will denote by  $|G|$  the number of vertices in  $G$  and by  $e(G)$  the number of edges in  $G$ . For  $v \in G$  we denote by  $N(v)$  the neighbourhood of  $v$  and by  $d(v) = |N(v)|$  its degree in  $G$ . For  $S \subseteq G$  we denote the induced subgraph of  $G$  on this subset by  $G[S]$ . Whenever working with graphs satisfying  $\alpha_m(G) \geq r$  all our asymptotics are with respect to  $n = |G|$  and we treat  $m$  and  $r$  as constants unless otherwise specified. When working with directed graphs  $N^\pm(v)$  denotes the in/out neighbourhood of  $v$  and  $d^\pm(v)$  in/out degree.

## 2 Lifting the Lower Bounds from Local to Global Independence Number

In this section, we prove our lower bounds on  $\alpha(G)$  for a graph  $G$  satisfying  $\alpha_m(G) \geq r$ . We begin with our lower bound result for  $k = \lceil \frac{m}{r-1} \rceil = 3$ , namely Proposition 1.1.

**Proposition 1.1** *Let  $m = 2r - 2 + t$  for  $1 \leq t \leq r - 1$ . Then any  $n$ -vertex graph  $G$  satisfying  $\alpha_m(G) \geq r$  has  $\alpha(G) \geq \Omega(n^{1-1/\ell})$ , where  $\ell = \lfloor \frac{r-1}{t} \rfloor + 1$ .*

**Proof** Let us first assume we can find a vertex disjoint collection of  $a = t\ell - (r - 1)$  cycles  $C_{2\ell-1}$  and  $t - a$  cycles  $C_{2\ell+1}$ . Their union makes a subgraph of  $G$  of order  $a(2\ell - 1) + (t - a)(2\ell + 1) = (2\ell + 1)t - 2a = 2r - 2 + t = m$  and has no independent set of size larger than  $a(\ell - 1) + (t - a)\ell = t\ell - a = r - 1$ . This gives us a contradiction to  $\alpha_m(G) \geq r$ , which means such a union does not exist in  $G$ .

If we can find  $a$  cycles  $C_{2\ell-1}$  in  $G$ , then the remainder of the graph can not contain  $t - a$  cycles  $C_{2\ell+1}$ , this means that by removing at most  $m$  vertices from  $G$  we can find a subgraph which is  $C_{2\ell+1}$ -free. If there are fewer than  $a$  cycles  $C_{2\ell-1}$  in  $G$ , then we find a subgraph, again with at least  $n - m$  vertices which is  $C_{2\ell-1}$ -free. In either case, a classical result from [15] (see also [28, 36] for slight improvements) on cycle-complete Ramsey numbers tells us there is an independent set of size at least  $\Omega((n - m)^{1-1/\ell}) = \Omega(n^{1-1/\ell})$ , as desired.  $\square$

We now show how to generalise any improvement made in the  $r = 3$  case to half of the range for any  $k$ . It will be convenient to denote by  $f(n, m, r)$  the smallest possible size of  $\alpha(G)$  in an  $n$  vertex graph with  $\alpha_m(G) \geq r$ .

**Lemma 2.1** *Let  $k = \lceil \frac{m}{r-1} \rceil$  and  $\ell = m - (k - 1)(r - 1)$ . Provided  $\ell \leq \frac{r-1}{2}$  we have  $f(n, m, r) \geq \min\{f(n - m, 2k - 1, 3), f(n - m, k - 1, 2)\}$ .*

**Proof** Let  $G$  be a graph on  $n$  vertices with  $\alpha_m(G) \geq r$ . Let us first assume that we can find a vertex disjoint union consisting of  $\ell$  subgraphs on  $2k - 1$  vertices, each having no independent set of size 3, and  $r - 1 - 2\ell$  copies of  $K_{k-1}$ . This union is a subgraph on  $\ell(2k - 1) + (r - 1 - 2\ell)(k - 1) = (r - 1)(k - 1) + \ell = m$  vertices which has no independent set larger than  $2\ell + r - 1 - 2\ell = r - 1$ , contradicting  $\alpha_m(G) \geq r$ .

If we can find  $\ell$  such subgraphs on  $2k - 1$  vertices this means that the remainder of the graph can not contain  $r - 1 - 2\ell$  copies of  $K_{k-1}$ . Removing our subgraphs and a maximal collection of  $K_{k-1}$ 's in the remainder we obtain a subgraph on at least  $n - m$  vertices which is  $K_{k-1}$ -free, or in other words has  $\alpha_{k-1} \geq 2$  implying  $f(n, m, r) \geq f(n - m, k - 1, 2)$ . If there are fewer than  $\ell$  such subgraphs we may remove a maximal collection and obtain a subgraph on at least  $n - m$  vertices which contains no subgraphs on  $2k - 1$  vertices without an independent set of size 3. In other words, our new subgraph has  $\alpha_{2k-1} \geq 3$  so  $f(n, m, r) \geq f(n - m, 2k - 1, 3)$  as claimed.  $\square$

Note that  $f(n, k - 1, 2) \geq \alpha$  is equivalent to  $R(k - 1, \alpha) \leq n$  implying that the best known bound is

$$f(n, k - 1, 2) \geq n^{\frac{1}{k-2}-o(1)}. \tag{1}$$

This represents a natural barrier for our results since it seems very likely that  $f(n, 2k - 1, 3) \geq f(n, k - 1, 2)$ . On the other hand, the results of [30] and [4] may be stated as  $f(n, m, r) \geq f(n - m, k, 2)$ . Therefore, obtaining a lower bound for  $f(n, 2k - 1, 3)$ , better than  $f(n, k, 2)$ , immediately improves their bound whenever the above lemma applies, i.e.  $\ell \leq \frac{r-1}{2}$ . Our result in the next section gives a bound which is halfway (in terms of exponents) between the above bounds coming from Ramsey numbers of  $K_{k-1}$  and  $K_k$  vs large independent set.

Fig. 1  $H_7$

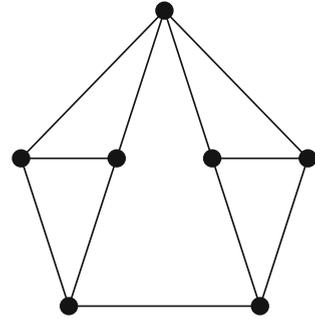
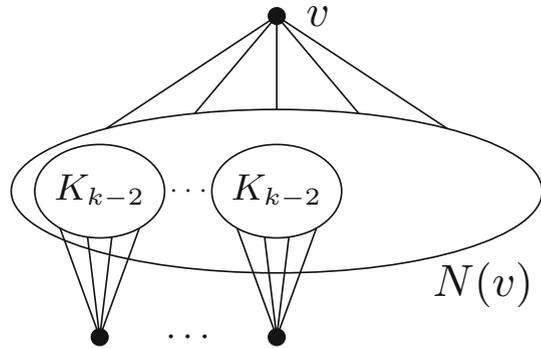


Fig. 2  $\mathcal{M}$



### 2.1 Independent Sets of Size Three Everywhere

In this subsection, we will show how to find big independent sets in graphs satisfying  $\alpha_{2k-1}(G) \geq 3$ . This is inherently a Ramsey question in the following sense. How big a graph do we need to take in order to guarantee that we can find an independent set of size  $\alpha$  or a subgraph  $H$  with  $2k - 1$  vertices and no independent set of size 3? The approach of [30] and [4] is to always look for a single graph  $H$ , namely a vertex disjoint union of  $K_k$  and  $K_{k-1}$ . This  $H$  clearly has no independent set of size 3 and the approach further reduces to finding a copy of  $K_k$ . The reason is that any bound strong enough to guarantee the existence of a  $K_k$  will remain strong enough to force a  $K_{k-1}$  once we remove the  $k$  vertices of a single copy of  $K_k$ , and thus force a copy of  $H$ . Our key new ingredient is to in addition look for a different  $2k - 1$ -vertex graph which we call  $H_{2k-1}$  and define to be a blow-up of  $C_5$  with parts of sizes  $1, k - 2, 1, 1, k - 2$  appearing in that order around the cycle, with cliques placed inside of parts (see Fig. 1 for an illustration). Since the complement of this graph is an actual blow-up of  $C_5$  it is triangle-free, implying that  $H_{2k-1}$  has no independent set of size 3 and is hence forbidden in any graph satisfying  $\alpha_{2k-1}(G) \geq 3$ .

We start by explaining the general idea behind our argument. As argued above our goal is to find a large independent set in an arbitrary  $K_k$ -free and  $H_{2k-1}$ -free  $n$ -vertex graph  $G$ . We will do so by finding a vertex  $v$  and a large collection of vertex disjoint

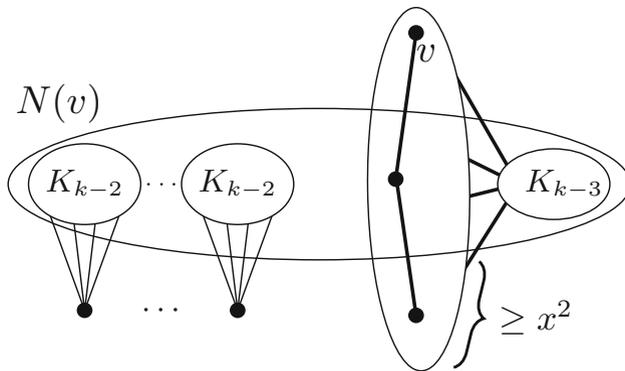


Fig. 3 Extending  $\mathcal{M}$

$K_{k-1}$ 's with  $k - 2$  vertices inside  $N(v)$ . We know that the remaining vertex of any such  $K_{k-1}$  lies outside  $N(v)$  as our graph is  $K_k$ -free. Furthermore, we know that the set of these last vertices spans an independent set as otherwise any edge between such vertices together with their  $K_{k-1}$ 's and  $v$  make a copy of  $H_{2k-1}$  (see Fig. 2 for an illustration). This gives us our desired large independent set.

The more difficult part of the argument is to actually find such a collection of  $K_{k-1}$ 's. The following two easy lemmas will help us control how many  $K_{k-1}$ 's we can find with  $k - 2$  vertices inside a neighbourhood of  $v$  and how many such  $K_{k-1}$ 's can intersect another one, respectively. Let us denote by  $t_i(G)$  the number of copies of  $K_i$  in  $G$ , (we omit  $G$  when it is clear from context).

**Lemma 2.2** *Let  $G$  be a graph with  $\alpha(G) \leq \alpha$  and let  $k \geq 2$ . Provided  $t_{k-1}(G) > 0$ , we have*

$$\frac{t_k(G)}{t_{k-1}(G)} \geq \frac{|G|}{k\alpha^{k-1}} - 1.$$

**Proof** We will prove the claim by induction on  $k$ . For the base case of  $k = 2$  (note that  $t_1 = |G|$ ) the claim follows from Turán's theorem which gives  $e(G) \geq \frac{n^2}{2\alpha} - \frac{n}{2}$  (see [3]). Let us now assume  $k \geq 3$  and that the claim holds for  $k - 1$ . Given  $S \subseteq V(G)$  let us denote by  $e(S)$  the number of edges with both endpoints in the common neighbourhood of  $S$  and by  $d(S)$  the number of common neighbours of  $S$ . Then we have

$$\begin{aligned} \binom{k}{2} t_k &= \sum_{G[S]=K_{k-2}} e(S) \geq \sum_{G[S]=K_{k-2}} \left( \frac{(d(S))^2}{2\alpha} - \frac{d(S)}{2} \right) \\ &\geq \frac{\left( \sum_{G[S]=K_{k-2}} d(S) \right)^2}{2t_{k-2}\alpha} - \sum_{G[S]=K_{k-2}} \frac{d(S)}{2} = \frac{(k-1)^2 t_{k-1}^2}{2t_{k-2}\alpha} - \frac{(k-1)t_{k-1}}{2} \end{aligned}$$

Where we used Turán's theorem within the common neighbourhood of  $S$  in the first inequality and the Cauchy–Schwarz inequality for the second. Dividing by

$(k - 1)t_{k-1}/2$  and using the induction assumption we obtain

$$\frac{kt_k}{t_{k-1}} \geq \frac{(k - 1)t_{k-1}}{\alpha t_{k-2}} - 1 \geq \frac{|G|}{\alpha^{k-1}} - \frac{k - 1}{\alpha} - 1 \geq \frac{|G|}{\alpha^{k-1}} - k.$$

□

**Lemma 2.3** *Let  $G$  be a  $K_k$ -free graph with  $\alpha(G) < \alpha$ . Then for any  $i \leq k$  we have*

$$t_i(G) \leq \frac{1}{i!} \cdot \alpha^{\binom{k}{2} - \binom{k-i}{2}}.$$

**Proof** We prove the claim by induction on  $i$ . For the base case of  $i = 1$  the claim is equivalent to  $t_1 = |G| \leq \alpha^{k-1}$  which holds by the classical bound on the Ramsey number  $R(k, \alpha)$ . Let us now assume the claim holds for  $i - 1$ . Given a subset of vertices  $S$  we denote by  $N(S)$  the set of common neighbours of  $S$  and by  $d(S) = |N(S)|$ . If  $G[S] = K_{i-1}$ , then  $N(S)$  is  $K_{k-i+1}$ -free in addition to having no independent set of size  $\alpha$ . So the same classical bound on Ramsey numbers as above implies  $d(S) \leq \alpha^{k-i}$ . Taking a sum and using the inductive assumption we obtain:

$$it_i = \sum_{G[S]=K_{i-1}} d(S) \leq \alpha^{k-i} t_{i-1} \leq \frac{1}{(i - 1)!} \cdot \alpha^{k-i + \binom{k}{2} - \binom{k-i+1}{2}} = \frac{1}{(i - 1)!} \cdot \alpha^{\binom{k}{2} - \binom{k-i}{2}}.$$

□

We are now ready to prove our general result for  $r = 3$ .

**Proof of Theorem 1.2** The main part of the proof will be to show that  $f(n, 2k - 1, 3) \geq \Omega(n^{1/(k-3/2)})$  but let us first verify this establishes the theorem in full generality. Indeed assuming this holds, given  $k = \lceil \frac{m}{r-1} \rceil$  and  $m \leq (k - \frac{1}{2})(r - 1)$  we have  $m - (k - 1)(r - 1) \leq \frac{r-1}{2}$  so that we may apply Lemma 2.1 giving us, when combined with (1), the desired result:

$$\begin{aligned} f(n, m, r) &\geq \min\{f(n - m, 2k - 1, 3), f(n - m, k - 1, 2)\} \\ &\geq \min \left\{ \Omega(n^{\frac{1}{k-3/2}}), n^{\frac{1}{k-2} - o(1)} \right\} \geq \Omega \left( n^{\frac{1}{k-3/2}} \right). \end{aligned}$$

Our remaining task now is to show that in an  $n$ -vertex graph  $G$  which contains an independent set of size 3 among any  $2k - 1$  vertices we can find an independent set of size  $\alpha = \Omega(n^{1/(k-3/2)})$ . With this choice of  $\alpha$ , we may assume that  $n \geq C\alpha^{k-3/2}$  for an arbitrarily large constant  $C$ . Since the result for  $k = 3$  holds by Proposition 1.1 we may assume  $k \geq 4$ .

As discussed above if  $G$  contains a  $K_k$ , then the remainder of the graph has no  $K_{k-1}$ , so by the classical Ramsey bound we get  $\alpha(G) \geq (n - k)^{1/(k-2)} \geq \Omega(n^{1/(k-3/2)})$ . So we may assume  $G$  is  $K_k$ -free. Our goal is to find a vertex  $v$  and a collection  $\mathcal{M}$  of  $\alpha$  vertex disjoint  $K_{k-1}$ 's each with  $k - 2$  vertices in  $N(v)$ . Then as we already explained above, any such  $K_{k-1}$  has exactly one vertex outside  $N(v)$ , as our graph is  $K_k$ -free. These vertices outside of  $N(v)$  form an independent set as otherwise any

edge between such vertices together with their  $K_{k-1}$ 's and  $v$  make a copy of  $H_{2k-1}$ , a contradiction. In order to do so we will analyse common neighbourhoods of cliques of size  $k - 3$  inside  $N(v)$ . Any edge we find inside such a common neighbourhood gives rise to a copy of  $K_{k-1}$  and if we find there a path of length 2 starting with  $v$ , then the last edge of this path gives rise to a copy of  $K_{k-1}$  with exactly  $k - 2$  vertices in  $N(v)$ , which we are looking for.

Let us first give a lower bound on  $T := (k - 2)t_{k-2}$  (recall that  $t_{k-2}$  denotes the number of  $K_{k-2}$ 's in  $G$ ) which counts extensions of a  $K_{k-3}$  into a  $K_{k-2}$ , i.e. the sum of sizes of common neighbourhoods of  $K_{k-3}$ 's in our graph. By a simple application of Lemma 2.2 we get:

$$T = (k - 2)t_{k-2} \geq t_{k-3} \cdot \left( \frac{n}{\alpha^{k-3}} - k + 2 \right) \geq t_{k-3} \cdot \frac{n}{2\alpha^{k-3}}, \tag{2}$$

where in the second inequality we used  $n \geq 2k\alpha^{k-3}$ , which holds for our choice of  $\alpha$ . Using Lemma 2.2 repeatedly in a similar way we get the following lower bound on  $T$  which will be useful later in the argument.

$$\begin{aligned} T &= (k - 2)t_{k-2} = (k - 2) \cdot \frac{t_{k-2}}{t_{k-3}} \cdots \frac{t_2}{t_1} \cdot n \\ &\geq (k - 2) \cdot \frac{n}{2(k - 2)\alpha^{k-3}} \cdots \frac{n}{4\alpha} \cdot n = \frac{n^{k-2}}{2^{k-3}(k - 3)! \alpha^{\binom{k-2}{2}}} \end{aligned} \tag{3}$$

In order to carry out the above proof strategy, for every clique of size  $k - 3$  we are going to restrict our attention only to a large part of its common neighbourhood where independent sets expand (meaning they have many vertices adjacent to some vertex of the set). This will achieve our goal since  $G$  being  $K_k$ -free means that the common neighbourhood of a  $K_{k-3}$  is triangle-free, so the part of this common neighbourhood inside  $N(v)$  is an independent set. Hence, expansion of this particular independent set precisely means there are many endpoints of a path of length 2 starting with  $v$ , giving many choices for a  $K_{k-1}$  with  $k - 2$  vertices in  $N(v)$ .

To obtain such an expansion we are going to first restrict attention to  $K_{k-3}$ 's which have the common neighbourhood of order at least half the average which equals  $T/t_{k-3}$ . We will call such a  $K_{k-3}$  *typical* and we know that altogether there are at least  $T/2$  ways of extending typical  $K_{k-3}$ 's into a  $K_{k-2}$ . By (2), any typical  $K_{k-3}$  has a common neighbourhood of size at least  $d := n/(4\alpha^{k-3})$ .

We now proceed to obtain independent set expansion inside the common neighbourhood of each typical  $K_{k-3}$ . Let  $S$  be such a  $K_{k-3}$ . Let  $X$  be a maximal independent set inside of the common neighbourhood  $N$  of  $S$  which has fewer than  $\frac{d-2\alpha}{2\alpha} \cdot |X|$  neighbours inside  $N$ , in case no such sets exists we set  $X = \emptyset$ . We now remove  $X$  and all its neighbours from  $N$ . Since  $X$  is an independent set we know  $|X| \leq \alpha$  so we have removed at most  $\alpha + \frac{d-2\alpha}{2\alpha} \cdot \alpha = \frac{d}{2}$  vertices from  $N$ . Any remaining vertex in  $N$  is called an *expanding neighbour* of  $S$ . Since  $S$  was arbitrary, every typical  $K_{k-3}$  has at least  $d/2$  expanding neighbours and inside its expanding neighbourhood independent sets expand by a factor of  $\frac{d-2\alpha}{2\alpha} \geq \frac{d}{4\alpha} = \frac{n}{16\alpha^{k-2}} =: d'$ . In particular, every vertex in  $N$  (being an independent set of size one) has degree at least  $d'$  in this set.

Furthermore, there are still at least  $T/4$  ways to extend a typical  $K_{k-3}$  into a  $K_{k-2}$  using an expanding neighbour.

We now pick our  $v$  to be a vertex which is an expanding neighbour of  $g_v \geq T/(4n)$  typical  $K_{k-3}$ 's (such  $v$  exists by double counting and the above bound). Let  $\mathcal{S}$  denote the collection consisting of all such  $K_{k-3}$ 's. Let us first observe some properties of an  $S \in \mathcal{S}$ . We denote by  $N_S$  its expanding neighbourhood and by  $D_S := N(v) \cap N_S$  the set of its expanding neighbours inside  $N(v)$ . By definition,  $v \in N_S$  for all  $S \in \mathcal{S}$ . Also, as was explained above, we know that  $v$  has at least  $d'$  neighbours within  $N_S$ , i.e.  $|D_S| \geq d'$ . These neighbours span an independent set (since they belong to the common neighbourhood of  $k - 2$  vertices in  $v \cup S$ ) of size at least  $d'$ , so they expand inside  $N_S$ . This gives us  $d'^2$  different vertices which together with  $S$  and one of the vertices in  $D_S$  make a  $K_{k-1}$  with exactly  $k - 2$  vertices inside  $N(v)$ . To find many such disjoint  $K_{k-1}$ 's we will use the fact that there are in total  $\sum_{S \in \mathcal{S}} |D_S| \geq g_v \cdot d'$  ways to extend  $K_{k-3}$ 's in  $\mathcal{S}$  into a  $K_{k-2}$  using an expanding neighbour belonging to  $N(v)$ .

Let us now consider a maximal collection  $\mathcal{M}$  of vertex disjoint  $K_{k-1}$ 's each with exactly  $k - 2$  vertices inside  $N(v)$ . Let us assume towards a contradiction that  $|\mathcal{M}| < \alpha$ . We will show below that if  $|\mathcal{M}| < \alpha$ , we can still find some  $S \in \mathcal{S}$  and a set  $D'_S \subseteq D_S$  of at least  $d'/2$  of its expanding neighbours in  $N(v)$  such that both  $S$  and  $D'_S$  are vertex disjoint from all cliques in  $\mathcal{M}$ . For now, suppose we found such  $S$  and  $D'_S$ . Since  $D'_S \subseteq D_S$  is an independent set (as we explained above), it expands within the neighbourhood of  $S$  meaning that there are at least  $d'^2/2 = \frac{n^2}{2^9 \alpha^{2k-4}} \geq \alpha$  vertices which together with  $S$  and some vertex in  $D'_S$  make a  $K_{k-1}$  with  $k - 2$  vertices in  $N(v)$ . Moreover, note that all these  $d'^2/2$  vertices lie outside of  $N(v)$  or we get a  $K_k$  in  $G$ . Since we removed fewer than  $\alpha$  vertices outside of  $N(v)$  (recall that each  $K_{k-1} \in \mathcal{M}$  has exactly one vertex outside  $N(v)$ ) one of these vertices is disjoint from all cliques in  $\mathcal{M}$  and gives rise to the desired copy of  $K_{k-1}$ , which is vertex disjoint from any clique in  $\mathcal{M}$  (since both  $S$  and  $D'_S$  are chosen disjoint from any clique in  $\mathcal{M}$ ) and hence contradicts its maximality.

Therefore, it remains to be shown that there is an  $S \in \mathcal{S}$  and  $d'/2$  of its expanding neighbours inside  $N(v)$ , all disjoint from any clique in  $\mathcal{M}$ . Note that cliques in  $\mathcal{M}$  cover at most  $\alpha(k - 2)$  vertices inside  $N(v)$ . On the other hand note that any vertex  $u \in N(v)$  can belong to at most  $\alpha \binom{k-2}{2} / (k - 3)!$  copies of  $K_{k-2}$  inside  $N(v)$ . This follows from Lemma 2.3 since any such copy of  $K_{k-2}$  amounts to a copy of  $K_{k-3}$  in the common neighbourhood of  $v$  and  $u$  which spans a  $K_{k-2}$ -free graph with no independent set of size  $\alpha$  (or we are done). On the other hand, any  $K_{k-2}$  can be an extension of at most  $k - 2$  different copies of  $K_{k-3}$  in  $\mathcal{S}$  so there are at least  $g_v d' - \alpha \binom{k-2}{2} + 1 (k - 2)^2 / (k - 3)!$  extensions of a  $K_{k-3}$  from  $S$  to a  $K_{k-2}$  inside  $N(v)$  both disjoint from any clique in  $\mathcal{M}$ . Note that

$$\begin{aligned} \frac{g_v d'}{\alpha \binom{k-2}{2} + 1 (k - 2)^2 / (k - 3)!} &\geq \frac{n^{k-3} / (2^{k-1} \alpha \binom{k-2}{2}) \cdot n / (16\alpha^{k-2})}{\alpha \binom{k-2}{2} + 1 (k - 2)^2} \\ &= \frac{n^{k-2}}{2^{k+3} (k - 2)^2 \alpha^{(k-2)^2+1}} \geq \frac{1}{2} \end{aligned}$$

where in the first inequality we used  $g_v \geq T/(4n)$  and (3) to bound  $T$ , while in the last inequality we used that  $n \geq C\alpha^{k-3/2} \geq C\alpha^{k-2+1/(k-2)}$ . This means that there are at least  $g_v d'/2$  such extensions and since  $g_v = |S|$  there must be a  $K_{k-3}$  with  $d'/2$  extensions, as desired.  $\square$

### 2.2 The Erdős–Hajnal (7,3) Case

For  $k = 4$  the result from the previous section implies that graphs with  $\alpha_7 \geq 3$  have  $\alpha \geq \Omega(n^{2/5})$  which already suffices to confirm the conjecture of Erdős and Hajnal [12]. In this section, we show how to further improve this bound to  $\alpha \geq n^{5/12-o(1)}$ , i.e. we prove Theorem 1.3.

The general idea will be similar as in the previous subsection. Here since  $k = 4$  we may assume our  $G$  satisfying  $\alpha_7(G) \geq 3$  is  $K_4$  and  $H_7$ -free. In fact, in this case,  $\alpha_7(G) \geq 3$  is essentially (up to removal of a few vertices) equivalent to  $G$  being  $K_4$  and  $H_7$ -free. Since we do not need the non-obvious direction here, we prove it as Lemma 3.6 in the following section where it will be useful.

Unlike in the previous section, since the desired  $\alpha$  is bigger we will not be able to find a large enough set of vertex disjoint triangles (not containing  $v$ ) with an edge in  $N(v)$ . However, these triangles will still play a major role in the argument. We will call them  $v$ -triangles and the vertex of a  $v$ -triangle not in  $N(v)$  is going to be called a  $v$ -extending vertex (in other words any non-neighbour of  $v$  which belongs to a  $K_4$  minus an edge together with  $v$ ). While we can not find a large enough collection of disjoint  $v$ -triangles the fact our graph is  $H_7$ -free imposes many restrictions on the subgraph induced by  $v$ -extending vertices, which we call  $N_\Delta(v)$ . The following lemma establishes the properties of  $N_\Delta(v)$  that we will use in our argument.

**Lemma 2.4** *Let  $G$  be a  $K_4$  and  $H_7$ -free graph and  $v \in G$ . Then*

- (a)  $N_\Delta(v) \cap N(v) = \emptyset$ .
- (b) *If  $u, w \in N_\Delta(v)$  belong to vertex disjoint  $v$ -triangles, then  $u \approx w$ .*
- (c)  $N_\Delta(v)$  is triangle-free.
- (d) *Let  $C$  be a connected subgraph of  $N_\Delta(v)$  consisting only of vertices belonging to at least 5 different  $v$ -triangles. If  $|C| \geq 2$ , then there exists  $u \in N(v)$  such that for any  $w \in C$ , all the edges within  $N(v) \cap N(w)$  make a star centred at  $u$  in  $G$ .*

**Proof** Part (a) is immediate since  $G$  is  $K_4$ -free and part (b) since it is  $H_7$ -free.

For part (c) assume to the contrary that there is a triangle  $x, y, z$  in  $N_\Delta(v)$  and that  $f, g, h$  are edges in  $N(v)$  completing a  $v$ -triangle with  $x, y, z$  respectively. By part (b) any two of  $f, g, h$  need to intersect. This is only possible if they make a triangle, in which case together with  $v$  they make a  $K_4$ , or if they make a star, in which case the centre of the star together with  $x, y, z$  makes a  $K_4$ , either way we obtain a contradiction.

For part (d) let  $u, w \in N_\Delta(v)$  be adjacent and belong to at least 5 different  $v$ -triangles. We claim that then edges within  $(N(w) \cap N(v)) \cup (N(u) \cap N(v))$  make a star in  $G$ . Indeed if  $N(u) \cap N(v)$  would contain two disjoint edges, then by part (b) any edge in  $N(w) \cap N(v)$ , of which there are at least 5 by assumption, must intersect them both. Since there can be at most 4 edges which intersect both of the two disjoint

edges, we conclude there can be no disjoint edges in  $N(u) \cap N(v)$ . This means that the edges span either a star or a triangle. Since there are at least 5 edges, the former must occur. We can repeat for  $w$  in place of  $u$  and observe that the only way for each pair of edges, one per star, to intersect is that they share the centre, as claimed. Propagating along any path in  $C$  we deduce that the same holds for any pair of vertices in  $C$ .  $\square$

The following corollary, based mostly on part d) of the above lemma, allows us to partition  $N_{\Delta}(v)$  into three parts which we will deal with separately in our argument.

**Corollary 2.5** *Let  $G$  be a  $K_4$  and  $H_7$ -free graph and  $v \in G$ . Then there exists a partition of  $N_{\Delta}(v)$  into three sets  $L$ ,  $I$  and  $C$  with the following properties:*

- (a)  $L$  consists only of vertices belonging to at most 4 different  $v$ -triangles.
- (b)  $I$  is an independent set.
- (c)  $C$  can be further partitioned into  $C_1, \dots, C_m$  such that there are no edges between different  $C_i$ 's and for every  $C_i$  there is a distinct  $v_i \in N(v)$  such that any  $v$ -triangle containing a vertex from  $C_i$  must contain  $v_i$  as well.

**Proof** We chose  $L$  to consist of all  $v$ -extending vertices belonging to at most 4  $v$ -triangles. We chose  $I$  to consist of isolated vertices in  $G[N_{\Delta}(v) \setminus L]$  and  $C = N_{\Delta}(v) \setminus (L \cup I)$ . Note that  $C$  is a union of connected components of  $G[N_{\Delta}(v) \setminus L]$ , which we denote by  $C'_1, C'_2, \dots$  each of order at least 2 and consisting entirely of vertices belonging to at least 5 different  $v$ -triangles. In particular, Lemma 2.4 part d) implies that there exists a vertex  $v'_i \in N(v)$  such that any  $v$ -triangle containing a vertex from  $C'_i$  must also contain  $v'_i$ . Finally, we merge any  $C'_i$ 's which have the same vertex for their  $v'_i$  to obtain the desired partition  $C_1, \dots, C_m$  of  $C$ .  $\square$

We are now ready to prove Theorem 1.3.

**Proof of Theorem 1.3** Let  $G$  be an  $n$ -vertex graph with  $\alpha_7(G) \geq 3$ . Our task is to show it has an independent set of size  $\alpha = n^{5/12 - o(1)}$ . If  $G$  contains a  $K_4$  the remainder of the graph must be triangle-free so has an independent set of size  $\Omega(\sqrt{n}) > \alpha$ . Hence, we may assume  $G$  is  $K_4$ -free as well as  $H_7$ -free.

We begin by ensuring the minimum degree is high, so that we can ensure good independent set expansion inside neighbourhoods. We repeatedly remove any vertex with degree at most  $2n/\alpha$ . Observe that if we remove more than half of the vertices, then the removed vertices induce a subgraph with at least  $n/2$  vertices and at most  $n \cdot 2n/\alpha$  edges. So, Turán's theorem implies there is an independent set of size at least  $\frac{n/2}{8n/\alpha+1} \geq \Omega(\alpha)$  and we are done. Let us hence assume that  $G$  has minimum degree at least  $2n/\alpha$  (we technically need to pass to a subgraph on at least  $n/2$  vertices but this only impacts the constants).

Let us fix a vertex  $v$ . Let  $X$  be a maximal independent set inside  $N(v)$  which has fewer than  $n/\alpha^2 \cdot |X|$  neighbours inside  $N(v)$ , and we set  $X = \emptyset$  if such a set does not exist. Since  $X$  is independent we may assume it has size at most  $\alpha/2$  so  $|X| + |N(X) \cap N(v)| \leq \alpha/2 + n/\alpha^2 \cdot \alpha/2 \leq n/\alpha$ . We direct<sup>1</sup> all edges from  $v$  towards

<sup>1</sup> The assignment of directions is simply a convenient way to encode the information about in what part of  $N(v)$  we know independent sets expand. An out-neighbour corresponds to an expanding neighbour in the previous argument.

$N(v) \setminus (X \cup N(X))$ . Note that  $d^+(v) \geq d(v)/2 \geq n/\alpha$  and inside  $N^+(v)$  independent sets expand by at least a factor of  $x := n/\alpha^2 \geq \Omega(n^{2/12})$ . We repeat for every vertex  $v$ , and note that some edges of  $G$  might be assigned both directions, while some other edges none.

Our first goal is to show there are in total many  $v$ -triangles for which their edge in  $N(v)$  belongs to  $N^-(v)$  and the remaining two edges of the triangle are directed away from this edge. We will call such a  $v$ -triangle a *directed  $v$ -triangle*. In other words, we are counting the number of  $K_4$ 's minus an edge with the 4 edges incident to the missing edge all being directed towards vertices of the missing edge. We denote this count by  $T_4$ .

**Claim** *Unless  $\alpha(G) \geq \Omega(n^{5/12})$  we have  $T_4 \geq \Omega(n^2)$ .*

**Proof** We will show that for any vertex  $v$  with in-degree at least half the average we get at least  $\Omega(d^-(v)^2/n^{2/12})$  directed  $v$ -triangles. Let us for now assume this holds. Note that  $v$  with lower in-degree contribute at most half to the value of  $\sum_{v \in G} d^-(v) = \sum_{v \in G} d^+(v) \geq n^2/\alpha \geq \Omega(n^{19/12})$  (recall that  $d^+(v) \geq n/\alpha$ ). An application of Cauchy–Schwarz implies there are at least  $\Omega((\sum_{v \in G} d^-(v))^2/n \cdot n^{-2/12}) \geq \Omega(n^2)$  directed triangles.

Let us now fix a vertex  $v$  with in-degree at least half the average, this in particular means  $d^-(v) \geq n/(2\alpha) \geq n^{7/12}$ . We know that any vertex  $u \in N^-(v)$  has  $v$  as an out-neighbour which means that  $v$ , as a single vertex independent set, expands inside  $N^+(u)$ . So  $v$  has at least  $x$  neighbours inside  $N^+(u)$ , i.e.  $u$  has at least  $x$  out-neighbours inside  $N(v)$ . Since  $u$  was an arbitrary vertex in  $N^-(v)$  this means there are at least  $d^-(v)x \geq \Omega(n^{9/12})$  edges inside  $N(v)$  which are directed away from a vertex in  $N^-(v)$  (note that if an edge of  $G$  has both directions and is inside of  $N^-(v)$ , then it is counted twice and considered as 2 directed edges). Let us call the set of such directed edges  $M$ , so in particular  $|M| \geq d^-(v)x \geq \Omega(n^{9/12})$ . Note further that given such a directed edge  $uw \in M$ , since  $u$  has  $w$  as a single vertex independent set in its out-neighbourhood  $w$  expands there. This means that our edge lies in at least  $x$   $v$ -triangles with both edges incident to  $u$  directed away from  $u$ . We call the third vertex of any such triangle  $uw$ -extending, note that it belongs to  $N_\Delta(v)$ . We are now going to assign types to edges in  $M$  according to where, inside  $N_\Delta(v)$ , we find the majority of their extending vertices.

Let us fix a partition of  $N_\Delta(v)$  into  $C$ ,  $L$  and  $I$ , provided by Corollary 2.5. Given a directed edge  $uw \in M$  we say it is of type  $L$  or  $C$  if it has at least  $x/3$  extending neighbours in  $L$  or  $C$ , respectively. We say it is of type  $I$  if it was not yet assigned a type. In particular, an edge of type  $I$  also has at least  $x/3$  extending vertices in  $I$  (since we have shown above it has at least  $x$  in total), but we also know it has at most  $2x/3$  extending vertices in other parts of  $N_\Delta(v)$ .

*First case:* at least a third of the edges in  $M$  are of type  $C$ .

Any vertex in  $N(v)$  is adjacent to fewer than  $\alpha$  other vertices inside  $N(v)$  (since these vertices make an independent set as  $G$  is  $K_4$ -free). This means there is a matching  $M'$  of at least  $|M|/(6\alpha)$  edges of type  $C$ , as otherwise, vertices making a maximal matching are incident to fewer than  $|M|/3$  edges so it can be extended. Let us denote by  $N_e$  the set of extending vertices of an edge  $e \in M'$  inside  $C$ . By assumption,  $|N_e| \geq x/3$  and note that each  $N_e$  spans an independent set (being in a common neighbourhood of  $e$ ).

On the other hand, we also claim  $N_e$ 's are disjoint. To see this suppose  $x \in N_e \cap N_{e'}$  for two distinct edges  $e, e' \in M'$ . By definition of  $N_e$ , there is some  $i$  such that  $x \in C_i$  and by Corollary 2.5 part c) we know that  $v_i \in e$  and  $v_i \in e'$ , a contradiction. Note also that there can be no edges between distinct  $N_e, N_{e'}$  or we find an  $H_7$ . This means that  $\bigcup_{e \in M'} N_e$  is an independent set of size at least  $x/3 \cdot |M|/(6\alpha) = \Omega(n^{6/12})$ .

*Second case:* at least a third of the edges in  $M$  are of type  $L$ .

Since any vertex in  $L$  belongs to at most 4 distinct  $v$ -triangles it can in particular be an extending vertex of at most 4 edges from  $M$ . Since every edge of type  $L$  has at least  $x/3$  extending vertices in  $L$  this means  $|L| \geq x|M|/12 \geq n^{11/12}/12$ . Now, Lemma 2.4 part c) implies  $L \subseteq N_\Delta(v)$  is triangle-free so there is an independent set of size at least  $\sqrt{|L|} \geq \Omega(n^{5.5/12})$ .

*Third case:* at least a third of the edges in  $M$  are of type  $I$ .

Let  $u \in N^-(v)$ . Let us denote by  $T_u$  the set of out-neighbours of  $u$  which together with  $u$  make an edge of type  $I$ . We know by the case assumption that  $\sum_{u \in N^-(v)} |T_u| \geq |M|/3$ . Note that  $T_u \subseteq N(v) \cap N^+(u)$  so must span an independent set (or we find a  $K_4$  in  $G$ ) and hence expands inside  $N^+(u)$ . This means there are at least  $x|T_u|$  distinct vertices extending an out-edge of  $u$  of type  $I$ . Note however that we do not know that all of them must be in  $I$ . But since any edge of type  $I$  has at most  $2x/3$  extending neighbours outside of  $I$  this means that there are least  $x|T_u|/3$  vertices in  $I$  extending an out-edge of  $u$ . This in particular means that  $u$  sends at least  $x|T_u|/3$  edges directed towards  $I$ . By taking the sum over all  $u$  we obtain that the number of edges directed from  $N^-(v)$  to  $I$  is at least  $xM/9 \geq d^-(v)x^2/9 \geq \Omega(n^{11/12})$ .

Since  $I$  spans an independent set we may assume  $|I| < \alpha$ . Let  $S_u = N^-(v) \cap N^-(u)$  for any  $u \in I$ . Let  $s_u = |S_u|$  so we know that  $\sum_{u \in I} s_u \geq d^-(v)x^2/9 \geq \Omega(n^{11/12})$ . Let  $I'$  be the subset of  $I$  consisting of vertices  $u$  with  $s_u \geq 2\alpha$ . Since vertices of  $I \setminus I'$  contribute at most  $2|I|\alpha < n^{10/12}$  to the above sum, we still have  $\sum_{u \in I'} s_u \geq d^-(v)x^2/18$ . By Turán's theorem for any  $u \in I'$  there needs to be at least  $s_u^2/(4\alpha)$  (using that  $s_u \geq 2\alpha$ ) edges inside  $S_u$ , or we find an independent set of size  $\alpha$ . Each such edge gives rise to a directed  $v$ -triangle. Hence, using Cauchy–Schwarz, there are at least  $\sum_{u \in I'} s_u^2/(4\alpha) \geq \Omega(d^-(v)^2x^4/\alpha^2) \geq \Omega(d^-(v)^2/n^{2/12})$  directed  $v$ -triangles, as desired. □

Let us give some intuition on how we are going to use the fact that  $T_4$  is big. Let us denote by  $\vec{d}_e$  the number of common out-neighbours of vertices making an edge  $e$ . Recall that  $T_4$  can be interpreted as the number of  $K_4$ 's minus an edge with all its edges oriented towards the missing edge. We call the remaining edge (for which we are not insisting on the direction) the *spine*. Note that the number of our  $K_4$ 's minus an edge having some fixed edge  $e$  as the spine is precisely  $\binom{\vec{d}_e}{2}$ . This means that  $T_4 = \sum_{e \in E(G)} \binom{\vec{d}_e}{2}$ . Our bound on  $T_4$  obtained above tells us that in a certain average sense the  $\vec{d}_e$ 's should be big.

Observe now that if one finds a star centred at  $v$  consisting of  $s$  edges each with  $\vec{d}_e \geq t$ , then this means that the  $s$  leaves each have  $t$  out-neighbours inside  $N(v)$  (note that we are disregarding the information that they are in fact inside  $N^+(v)$ ). This will allow us to play a similar game as we did in the previous claim, indeed there we tackled the same problem with  $s = d^-(v)$  and  $t = x$  (which we obtained through expansion)

with an important difference, namely that the star was in-directed. The bound on  $T_4$  tells us that there is such a star with larger (in certain sense) parameters  $s$  and  $t$  and the next claim shows how this gives rise to many of our  $K_4$ 's minus an edge, which we find in a different place (in particular they do not use the centre of the star since we do not know the direction of centre's edges).

For any  $v \in G$  let  $S_v$  be the star consisting of a centre  $v$  and all its edges in  $G$ . Let us also denote by  $s_v$  the total number of our  $K_4$ 's minus an edge with an edge of  $S_v$  as their spine, i.e.  $s_v = \sum_{e \in S_v} \binom{\vec{d}_e}{2}$ . Summing over  $v$  we also have  $2T_4 = \sum_v \sum_{e \in S_v} \binom{\vec{d}_e}{2} = \sum_v s_v$ .

**Claim** Unless  $\alpha(G) \geq n^{5/12-o(1)}$  and provided  $s_v \geq \frac{T_4}{n}$  there exist  $\frac{s_v^2}{n^{7/12+o(1)}}$  of our  $K_4$ 's minus an edge with their spine inside  $N(v)$ .

**Proof** Let us first “regularise”  $\vec{d}_e$ 's for  $e \in S_v$ . Let us partition these edges into at most  $\log n$  sets with all edges belonging to a single set having  $\vec{d}_e \in [t, 2t]$  for some  $t$ . Since  $s_v = \sum_{e \in S_v} \binom{\vec{d}_e}{2}$  and the edges in  $S_v$  are split into at most  $\log n$  sets, we conclude that one set contributes at least  $\frac{s_v}{\log n}$  to this sum. I.e. edges in this set make a substar  $S$  of  $S_v$  consisting of  $s$  edges, each having  $t \leq \vec{d}_e \leq 2t$  for some  $s, t$  satisfying  $2st^2 \geq s \binom{2t}{2} \geq \frac{s_v}{\log n} \geq \frac{T_4}{n \log n} \geq n^{1-o(1)}$ .<sup>2</sup> We may assume that  $t \leq n^{5/12}$  since  $\vec{d}_e$  counts certain common neighbours of a fixed edge which must span an independent set (or there is a  $K_4$ ).

Similarly, as in the previous claim, we define  $M$  to be the set of directed edges inside  $N(v)$  with source vertex being a leaf of  $S$ . Now the fact that any edge  $e \in S$  has  $\vec{d}_e \geq t$  means that any leaf  $u$  of  $S$  is a source of at least  $t$  edges in  $M$ . Let us remove all but exactly  $t$  such edges from  $M$ , so, in particular,  $|M| = st$  (as before while some edges might be oriented both ways we treat this as two distinct directed edges). Let us denote by  $T_u$  the set of  $t$  out-neighbours of  $u$  which together with  $u$  make an edge in  $M$ . We know  $T_u$  is an independent set (it consists of common neighbours of the edge  $vu$ ). In particular, both  $T_u$  and any of its subsets expand inside  $N^+(u)$ . Let us consider an auxiliary bipartite graph with the left part being  $T_u$  and the right part being  $N_\Delta(v) \cap N^+(u)$ . We put an edge between two vertices if together with  $u$  they make a triangle in  $G$ . The expansion property translates to the fact that any subset of size  $t'$  of the left part has at least  $t'x$  distinct neighbours on the right. A standard application of Hall's theorem (to the graph obtained by taking  $x$  copies of every vertex on the left) tells us we can find  $t$  disjoint stars each of size  $x$  in this graph. Translating back to our graph, for each of the  $t$  edges incident to  $u$  in  $M$  we have found a set of  $x$  out-neighbours of  $u$  which extend it into a triangle. Moreover, these sets are disjoint for distinct edges. We call these  $x$  vertices extending for the corresponding edge.<sup>3</sup>

Once again let us take a  $C, L, I$  partition provided by Corollary 2.5 and assign types  $C, L$  and  $I$  to edges in  $M$ , which have at least  $x/3$  extending neighbours in  $C, L$

<sup>2</sup> We decided to pay the  $\log n$  factor here for simplicity, it is possible to do the same argument more carefully and avoid it.

<sup>3</sup> Note that this is a subset of what we considered to be extending vertices in the previous claim. Here it is important for us to fix the number of extending neighbours for every edge for certain regularity considerations.

and  $I$ , respectively. This is similar to the previous argument except that we are using our new, slightly modified definition of extending vertices. We again split into three cases according to which type is in the majority.

*First case:* at least  $st/3$  edges in  $M$  are of type  $C$ .

As in 1. case of the previous claim we can find a matching  $\mathcal{M}$  of at least  $st/(6\alpha)$  edges of type  $C$  (since vertices of  $\mathcal{M}$  are incident to at most  $2|\mathcal{M}|\alpha$  edges inside  $N(v)$ ). Let us denote by  $N_e$  the set of extending neighbours of an edge  $e \in \mathcal{M}$  inside  $C$ , so  $|N_e| \geq x/3$ . Again, as before, each  $N_e$  is independent (neighbours of the same edge), and they are disjoint for different  $e$  (otherwise if a vertex belonging to two  $N_e$ 's belongs to  $C_i$ , we get a contradiction to the uniqueness of  $v_i$ ) and there are no edges between distinct  $N_e$ 's (or we find an  $H_7$ ). So their union makes an independent set of size at least  $stx/(18\alpha)$ . If  $t \leq n^{4/12}$ , then  $st^2 \geq n^{1-o(1)}$  implies  $st \geq n^{8/12-o(1)}$  and our independent set is of size at least  $n^{5/12-o(1)}$ , as desired. So let us assume  $t \geq n^{4/12}$ .

Note that by Corollary 2.5, given an edge  $e \in M$  and its extending neighbour which belongs to some  $C_i$  we know that  $v_i \in e$  and  $v_i$  can be either source or sink of  $e$ . In the former case, we say the neighbour is source-extending and in the latter sink-extending. We say  $e$  is of type  $C$ -source if it has at least  $x/6$  source-extending neighbours in  $C$  and of type  $C$ -sink if it has at least  $x/6$  sink-extending neighbours in  $C$ .

If there are at least  $st/6$  edges in  $M$  of type  $C$ -sink that means there is a leaf  $u$  of  $S$  which is a startpoint of at least  $t/6$  such edges. If  $uw$  is one of these  $t/6$  edges it has a set of at least  $x/6$  sink-extending neighbours which span an independent set (being common neighbours of an edge) and all belong to the same  $C_i$  (namely the one for which  $v_i = w$ ) and these  $C_i$ 's are distinct between edges. This means that these sets are disjoint between ones corresponding to distinct edges and span an independent set (there are no edges between distinct  $C_i$ 's) so we found an independent set of size at least  $tx/36 \geq \Omega(n^{6/12})$ .

If there are at least  $st/6$  edges in  $M$  of type  $C$ -source we need to be able to find  $s/12$  leaves of  $S$  incident to at least  $t/12$  such edges each (since we know that every leaf of  $S$  is incident to exactly  $t$  edges in  $M$  so otherwise, there would be fewer than  $s/12 \cdot t + s \cdot t/12 = st/6$  such edges in total). For any such leaf  $u$  this means we find  $\frac{t}{12} \cdot \frac{x}{6} \geq \frac{tx}{72}$  source-extending neighbours of its edges (using our preprocessing fact that extending neighbours of distinct edges incident to  $u$  are disjoint). By Lemma 2.4 c) we know there is an independent set of size  $\sqrt{tx}/9$  among these neighbours. Since these are source-extending neighbours we know they all belong to  $C_i$  for which  $v_i = u$ . In particular, for distinct  $u$  they belong to distinct  $C_i$ 's, meaning we obtain an independent set of size  $\Omega(s\sqrt{tx}) = \Omega(st^2\sqrt{x}/t^{3/2}) \geq n^{13/12-o(1)}/t^{3/2} \geq n^{5.5/12-o(1)}$ , using  $t \leq n^{5/12}$ .

*Second case:* at least  $st/3$  edges in  $M$  are of type  $L$ .

Let us first assume  $s \geq t$ . We again find a matching of size  $st/(6\alpha)$  of edges of type  $L$  in  $M$ . Each edge  $e$  in the matching gives rise to a set  $N_e$  of  $x/3$  extending neighbours in  $L$ .  $N_e$  spans an independent set (being inside the common neighbourhood of an edge) and there can be no edges between distinct  $N_e$ 's (or we find an  $H_7$ ). This means that the union of  $N_e$ 's spans an independent set. Since any extending neighbour in this union can be extending for at most 4 edges (so belongs to at most 4 different  $N_e$ 's) this gives  $\alpha(G) \geq stx/(72\alpha) \geq (st^2)^{2/3}x/(72\alpha) \geq n^{5/12-o(1)}$ , using  $s \geq t$  and  $st^2 \geq n^{1-o(1)}$ .

Let us now assume  $t \geq s$ . We can find at least  $s/6$  leaves of  $S$  each being a start vertex of at least  $t/6$  edges in  $M$  of type  $L$  (otherwise there would be fewer than  $s/6 \cdot t + s \cdot t/6 = st/3$  edges in total). Given a directed edge  $uw \in M$  of type  $L$ , with  $u$  being one of these  $s/6$  leaves, we define  $A_{uw}$  as the set of extending vertices of  $uw$  belonging to  $L$ . We will now state some properties of these sets  $A_{uw}$ , which will allow us to find a big independent set. Since this is the most technical part of the proof and, once the appropriate properties are identified, is independent of the rest of the argument we prove it as a separate lemma afterwards.

- (1) No vertex belongs to more than 4 different  $A_{uw}$ 's. Since  $A_{uw} \subseteq L$  and by Corollary 2.5 part (a), any vertex in  $L$  belongs to at most 4 different  $v$ -triangles, this means it can belong to at most 4 different  $A_{uw}$ 's.
- (2)  $|A_{uw}| \leq x$ . This follows since, by our definition, there are exactly  $x$   $uw$ -extending vertices.
- (3)  $\sum_w |A_{uw}| \geq tx/18$ . This follows since for any  $u$  there are at least  $t/6$  edges  $uw$  for which  $A_{uw}$  is defined and each such edge being of type  $L$  means there are at least  $x/3$  extending vertices, meaning  $|A_{uw}| \geq x/3$ .
- (4) If  $uw$  and  $u'w'$  are independent, then there can be no edges between  $A_{uw}$  and  $A_{u'w'}$ . Else, we find  $H_7$ .

This precisely establishes the conditions of Lemma 2.6, which provides us with an independent set of size  $\min(\Omega(s\sqrt{tx}), \Omega(s^{1/2}t^{3/4}x^{1/4}), \Omega(s^{3/5}t^{3/5}x^{2/5}))$ . Each of the three expressions is minimised when  $t$  is as large as possible (under the assumption  $st^2 \geq n^{1-o(1)}$ ) so we may plug in  $t = n^{5/12}$  and  $s = n^{2/12-o(1)}$  in which case the first expression evaluates to  $n^{5.5/12-o(1)}$ , the second to  $n^{5.25/12-o(1)}$  and the third to  $n^{5/12-o(1)}$ .

*Third case:* there are at least  $st/3$  edges in  $M$  of type  $I$ .

Since  $I$  spans an independent set we know  $|I| < \alpha$ . Any directed edge  $uw$  in  $M$  of type  $I$  has at least  $x/3$  extending neighbours in  $I$ . Since these are distinct for different  $w$ 's by our definition of an extending neighbour and since we insist that extending neighbours are out-neighbours of  $u$  this means that overall there are at least  $stx/9$  edges directed from leaves of  $S$  to  $I$ .

This means that the average out-degree from  $S$  to  $I$  is at least  $tx/9$ . This together with a standard application of Cauchy–Schwarz implies there are  $\Omega(s(tx)^2) = n^{16/12-o(1)}$  (recall that  $st^2 \geq s_v/(2 \log n) \geq n^{1-o(1)}$ ) out-directed cherries ( $K_{1,2}$ 's) with the centre in  $S$ . If we denote by  $\mathcal{P}$  the set of pairs of vertices in  $I$ , then  $|\mathcal{P}| = \binom{|I|}{2} \leq \alpha^2$ , let us also denote by  $d_p$  the number of common in-neighbours of a pair of vertices  $p \in \mathcal{P}$ . So in particular,  $\sum_{p \in \mathcal{P}} d_p = \Omega(s(tx)^2) = n^{16/12-o(1)}$ . Pairs  $p$  with  $d_p < 2\alpha$  contribute at most  $|\mathcal{P}| \cdot 2\alpha \leq n^{15/12-o(1)}$  to this sum so if  $\mathcal{P}' \subseteq \mathcal{P}$  denotes the set of pairs which have  $d_p \geq 2\alpha$ , then also  $\sum_{p \in \mathcal{P}'} d_p = \Omega(s(tx)^2) = n^{16/12-o(1)}$ . Applying Turán's theorem inside a common neighbourhood of  $p \in \mathcal{P}'$  we find there  $d_p^2/(4\alpha)$  edges (using that  $d_p \geq 2\alpha$ ) or there is an independent set of size  $\alpha$ . Note that any edge we find inside this common in-neighbourhood gives rise to our desired  $K_4$  minus an edge.

In particular, using Cauchy–Schwarz we find at least  $\sum_{p \in \mathcal{P}'} \frac{d_p^2}{4\alpha} \geq \frac{(\sum_{p \in \mathcal{P}'} d_p)^2}{4\alpha|\mathcal{P}'|} \geq \Omega(s^2(tx)^4/\alpha^3) \geq s_v^2/n^{7/12+o(1)}$  copies of our desired  $K_4$  minus an edge, as claimed. □

Recall that  $2T_4 = \sum_v s_v$ . Note also that stars with  $s_v \geq T_4/n$  contribute at least  $T_4$  to this sum. Hence, taking the sum over  $v$  of the number of copies of our  $K_4$ 's minus an edge with the spine in  $N(v)$  we obtain at least  $\sum s_v^2/n^{7/12+o(1)} \geq T_4^2/n^{19/12+o(1)} \geq T_4 n^{5/12-o(1)}$ , where we used Cauchy–Schwarz in the first inequality and our bound  $T_4 \geq \Omega(n^2)$ , from the first claim in the second. Note however that certain copies of our  $K_4$ 's minus an edge got counted multiple times. But, for every  $v$  that counted our  $K_4$  minus an edge, we know it had its spine inside  $N(v)$ . This means that a single copy could be counted at most  $\alpha(G)$  times since the spine (being an edge in  $G$ ) can have at most  $\alpha(G)$  neighbours, as they span an independent set. In particular, unless  $\alpha(G) \geq n^{5/12-o(1)}$ , this shows that there are more than  $T_4$  distinct copies of our  $K_4$ 's minus an edge, contradicting the definition of  $T_4$  and completing the proof.  $\square$

We now prove the lemma we used in the proof above. Let us first attempt to help the reader parse the statement. It says that if we can partition vertices of  $G$  into a grid of subsets each of size at most  $x$  (so each cell of the grid contains at most  $x$  vertices), such that  $G$  only has edges between vertices in the same row or column of the grid, and we additionally know that there is a large number of vertices in each row, then we can find a big independent set in the whole graph.

**Lemma 2.6** *Let  $G$  be a triangle-free graph with vertex set  $\bigcup_{i,j} A_{ij}$  where  $i \in [s]$ ,  $j \in \mathbb{N}$ . If*

1. *no vertex appears in more than 4 different  $A_{ij}$ 's;*
2.  *$|A_{ij}| \leq x$ , for any  $i, j$ ;*
3. *for some  $t \geq s$  and any  $i \in [s]$  there are at least  $tx$  vertices in  $\bigcup_j A_{ij}$ ;*
4. *there are no edges of  $G$  between  $A_{ij}$  and  $A_{k\ell}$  for any  $i \neq k$  and  $j \neq \ell$ ,*

*then  $\alpha(G) \geq \min(\Omega(s\sqrt{tx}), \Omega(s^{1/2}t^{3/4}x^{1/4}), \Omega(s^{3/5}t^{3/5}x^{2/5}))$ .*

**Proof** Let us first replace any vertex which appears in multiple  $A_{ij}$ 's with distinct copies of itself, one per  $A_{ij}$  it appears in. Our new graph has all  $A_{ij}$  disjoint and satisfies the same conditions as the original. In addition, the independence number went up by at most a factor of 4 so showing the result for our new graph implies it for the original. So let us assume that sets  $A_{ij}$  actually partition the vertex set of  $G$ .

Let  $\alpha = \alpha(G)$ . We will call  $\bigcup_j A_{ij}$  a row of our grid,  $\bigcup_i A_{ij}$  a column and each  $A_{ij}$  a cell. Let us first clean up the graph a bit. As long as we can find an independent set  $I$  of size more than  $2\alpha/s$  using vertices from at most  $t/s$  cells inside some row, we take  $I$ , delete the rest of the row and all the columns containing a vertex of  $I$  from  $G$ . If we repeat this at least  $s/2$  many times we obtain an independent set of size larger than  $\alpha$  which is impossible. This means that upon deleting at most  $s/2$  many rows and at most  $(t/s) \cdot s/2 = t/2$  many columns we obtain a subgraph for which in any row any  $t/s$  cells don't contain an independent set of size at least  $2\alpha/s$ . This subgraph still satisfies all the conditions of the lemma with  $t := t/2$  and  $s := s/2$ . The only non-immediate condition is 3, it holds since we deleted at most  $t/2$  cells in any of the remaining rows, so in total at most  $tx/2$  vertices in that row altogether, using that any cell contains at most  $x$  vertices. From now on we assume our graph  $G$  satisfies the property that in any row any  $t/s$  cells don't contain an independent set of size  $2\alpha/s$ .

Let us delete vertices from our graph until we have exactly  $tx$  in every row. Let  $n$  denote the number of vertices of  $G$ , so  $n = stx$ . Observe that at least half of the

vertices of  $G$  have degree at least  $n/(4\alpha)$  as otherwise, vertices with degree lower than this induce a subgraph which has an independent set of size at least  $\alpha$  by Turán’s theorem. Condition 4 ensures each such vertex either has at least  $n/(8\alpha)$  neighbours in its row or  $n/(8\alpha)$  neighbours in its column. In particular, at least a quarter of vertices of  $G$  fall under one of these cases. Since  $G$  is triangle-free, the neighbourhood of any vertex is an independent set. We conclude that either there are at least  $s/4$  rows containing an independent set of size  $n/(8\alpha)$  or there are  $t/4$  columns containing an independent set of size  $n/(8\alpha)$ .

Let us first consider the latter case. Let  $U$  be the union of our  $t/4$  independent sets of size at least  $n/(8\alpha)$ , belonging to distinct columns, so consisting of at least  $tn/(32\alpha)$  vertices. Let  $a_i$  denote the number of vertices of  $U$  in row  $i$ . Then  $\sum_{i=1}^s a_i \geq tn/(32\alpha)$  and each  $a_i \leq tx$  (since we removed all but  $tx$  vertices in any row). On the other hand, since  $G$  is triangle-free, we know that in each row we can find an independent subset of  $U$  of size  $\sqrt{a_i}$ . In particular, since  $U$  was constructed as a union of independent sets in columns (and all edges of  $G$  are either within columns or within rows) this means that  $U$  contains an independent set of size  $\sum_{i=1}^s \sqrt{a_i} \geq \frac{tn}{32\alpha \cdot tx} \cdot \sqrt{tx} = \Omega(st^{3/2}x^{1/2}/\alpha)$  (where we used  $n = stx$  and the standard fact that sum of roots is minimised, subject to constant sum, when as many terms as possible are as large as possible). In other words, we showed  $\alpha \geq \Omega(st^{3/2}x^{1/2}/\alpha)$  giving us the second term of the minimum.

Moving to the former case let us again take a union  $U$  of our  $s/4$  independent sets of size at least  $n/(8\alpha)$ , belonging to distinct rows, so again  $|U| \geq sn/(32\alpha)$ . Call a cell  $A_{ij}$  full if it contains at least  $4\alpha/t$  vertices of  $U$ . There are fewer than  $t/(2s)$  full cells in any row since otherwise,  $U$  restricted to  $\lceil t/(2s) \rceil$  full cells gives us an independent set of size at least  $2\alpha/s$  using at most  $\lceil t/(2s) \rceil \leq t/s$  cells (using  $t \geq s$ ), which contradicts our property from the beginning. Using this and once again the property from the beginning we conclude there can be at most  $2\alpha/s$  vertices of  $U$  in full cells of any fixed row. If  $2\alpha/s \geq n/(16\alpha)$ , then  $\alpha^2 \geq \Omega(ns) \geq \Omega(s^2tx)$ , so first term of the minimum is satisfied. So we may assume  $2\alpha/s \leq n/(16\alpha)$ . Hence, by removing from  $U$  any vertex belonging to a full cell we remove at most half the vertices of  $U$  (since  $U$  had at least  $n/(8\alpha)$  vertices in every row). Now, finally, denote by  $a_i$  the number of vertices of  $U$  belonging to the column  $i$ . So  $\sum_{i=1}^s a_i = |U| \geq sn/(64\alpha)$  and  $a_i \leq s \cdot 4\alpha/t$ , since all the remaining vertices of  $U$  belong to non-full cell. As before  $\alpha \geq \sum_{i=1}^s \sqrt{a_i} \geq \frac{sn/(64\alpha)}{4s\alpha/t} \cdot \sqrt{4s\alpha/t} \geq \Omega(s^{3/2}t^{3/2}x/\alpha^{3/2})$  giving us the third term of the minimum. □

### 3 2-Density and Local Independence Number

In this section, we show our upper bounds on the minimum possible  $\alpha(G)$  in a graph  $G$  satisfying  $\alpha_m(G) \geq r$ . To do this we need to exhibit a graph with no large independent set in which any  $m$ -vertex subgraph contains an independent set of size  $r$ . As discussed in the introduction the natural candidates are random graphs and the answer is controlled by  $M(m, r)$  which is defined to be the minimum value of the 2-density over all graphs  $H$  on  $m$  vertices having  $\alpha(H) \leq r - 1$ . It will be convenient to define  $d_2(H) = \frac{e(H)-1}{|H|-2}$ , so that the 2-density is defined as the maximum of  $d_2(H')$  over

subgraphs  $H'$  of order at least 3, from now on whenever we consider 2-density we will implicitly assume the subgraphs we take to have at least 3 vertices. We begin by proving our reduction to the 2-density Turán problem, namely Proposition 1.4.

**Proposition 1.4** *Let  $m, r$  be fixed,  $m \geq 2r - 1 \geq 3$  and  $M = M(m, r)$ . Then for any  $n$  there exists an  $n$ -vertex graph  $G$  with  $\alpha_m(G) \geq r$  and  $\alpha(G) \leq n^{1/M + o(1)}$ .*

**Proof** A graph  $H$  is said to be *strictly 2-balanced* if  $m_2(H) > m_2(H')$  for any proper subgraph  $H'$  of  $H$ . I.e., if  $H$  itself is the maximiser of  $d_2(H)$  among its subgraphs and in particular  $m_2(H) = d_2(H) = \frac{e(H)-1}{|H|-2}$ .

Let  $\mathcal{H} = \{H_1, \dots, H_t\}$  be a collection of strictly 2-balanced graphs such that any  $m$ -vertex  $H$  with  $\alpha(H) \leq r - 1$  contains some  $H_i$  as a subgraph. We can trivially obtain it by replacing any  $H$  in our family which is not strictly 2-balanced by its subgraph  $H'$  which maximises  $d_2(H')$ . In particular,  $\mathcal{H}$  is a family of strictly balanced 2-graphs  $H$  satisfying  $m_2(H) = d_2(H) \geq M$ , with at most  $m$  vertices and with the property that if a graph is  $\mathcal{H}$ -free, then it satisfies  $\alpha_m \geq r$ . Note that  $t \leq 2^{\binom{m}{2}}$ , so is in particular bounded by a constant (depending on  $m$ ).

Let  $G \sim \mathcal{G}(n, p)$ , where we choose  $p := 1/(48tn^{1/M})$  and will be assuming  $n$  to be large enough throughout. Let  $A_K$  denote the event that a subset  $K \subseteq V(G)$ , consisting of  $k := \frac{8 \log n}{p} + 2 = O(n^{1/M} \log n)$  vertices, spans an independent set. In particular, we have  $\mathbb{P}(A_k) = (1-p)^{\binom{k}{2}}$ . Let  $B_j^i$  denote the event that we find a copy of  $H_i$  at the  $j$ -th possible location (so fixing the subset of vertices of  $G$  where we could find  $H_i$  and the labellings of vertices). In particular,  $\mathbb{P}(B_j^i) = p^{e(H_i)}$ . Our goal is to show that with positive probability none of the events  $A_K$  or  $B_j^i$  occur, which implies that there is an  $\mathcal{H}$ -free graph with no independent set of size  $O(n^{1/M} \log n)$  as desired. We will do so by using the asymmetric version of the Lovász local lemma (see Lemma 5.1.1 in [3]). To apply the lemma we first need to understand how many dependencies there are between different types of events. In particular, given  $A_K$  it depends only on  $\binom{k}{2}$  edges of  $G$ , so in particular it is mutually independent of all events  $B_j^i$  which do not contain one of these edges. In particular, it is mutually independent from all but at most  $k^2 n^{|H_i|-2}$  events  $B_j^i$  and at most  $n^k$  other events  $A_{K'}$  (since there are at most this many such events in total). Similarly, any  $B_j^i$  is mutually independent of all but at most  $e(H_i)n^{|H_{i'}|-2} \leq m^2 n^{|H_{i'}|-2}$  events  $B_{j'}^{i'}$  for a fixed  $i'$  and at most  $n^k$  events  $A_k$ .

We now need to choose parameters  $x$  (corresponding to events of type  $A_K$ ) and  $y_i$  (corresponding to events of type  $B_j^i$ ) such that

$$\mathbb{P}(A_K) \leq x \cdot (1-x)^{n^k} \cdot \prod_i (1-y_i)^{k^2 n^{|H_i|-2}} \quad \text{and}$$

$$\mathbb{P}(B_j^i) \leq y_i \cdot (1-x)^{n^k} \cdot \prod_{i'} (1-y_{i'})^{m^2 n^{|H_{i'}|-2}}$$

which will complete the proof. We choose  $x = 1/n^k$  so that in particular  $(1-x)^{n^k} \geq 1/3$  (as  $n$  is large) and  $y_i = p/(8tn^{|H_i|-2}) \leq 1/2$  so that in particular  $(1-y_i)^{n^{|H_i|-2}} \geq$

$e^{-p/(4t)}$  (using  $1 - a \geq e^{-2a}$  for  $a \leq 1/2$ ). With these choices we obtain

$$\begin{aligned} & x \cdot (1 - x)^{n^k} \cdot \prod_i (1 - y_i)^{k^2 n^{|H_i|-2}} \\ & \geq \frac{1}{n^k} \cdot \frac{1}{3} \cdot e^{-pk^2/4} \geq e^{-2k \log n - pk^2/4} = e^{-p \binom{k}{2}} \geq (1 - p) \binom{k}{2} = \mathbb{P}(A_K) \quad \text{and} \\ & y_i \cdot (1 - x)^{n^k} \cdot \prod_{i'} (1 - y_{i'})^{m^2 n^{|H_{i'}|-2}} \\ & \geq \frac{p}{8tn^{|H_i|-2}} \cdot \frac{1}{3} \cdot e^{-m^2 p/4} \geq \frac{p}{48t \cdot n^{\frac{|H_i|-2}{e(H_i)-1} \cdot (e(H_i)-1)}} \geq \frac{p}{(48tn)^{\frac{1}{M} \cdot (e(H_i)-1)}} \\ & \geq p^{e(H_i)} \geq \mathbb{P}(B_i^j). \end{aligned}$$

Here, in the second inequality, we used the fact that  $m$  is a constant while  $p \rightarrow 0$  so  $m^2 p \rightarrow 0$  and in particular  $e^{-m^2 p/4} \leq 1/2$  (since  $n$  is large). In the third inequality we used  $n^{-\frac{|H_i|-2}{e(H_i)-1}} \geq n^{-\frac{1}{M}}$ , which follows since  $M$  is equal to the minimum of  $\frac{e(H_i)-1}{|H_i|-2}$  over  $H_i$  (and we used  $|H_i| \geq 3$  to put the  $48t$  factor under the exponent).  $\square$

**Remark** This result appears to be the best one can expect to get using random graphs, up to the polylog factor. The polylog factor can likely be slightly improved compared to the above argument by using the  $\mathcal{H}$ -free process (see e.g. [32] for more details about this process).

If we replace  $m_2(H)$  in the definition of  $M(m, r)$  with  $d_2(H) = \frac{e(H)-1}{|H|-2}$  the problem of determining  $M(m, r)$  would reduce to the classical Turán’s theorem. Indeed, since the number of vertices is fixed, minimising  $d_2(H)$  is tantamount to minimising the number of edges in an  $m$ -vertex graph with  $\alpha(H) < r$  and upon taking complements we reach the setting of the classical Turán’s theorem. This is why it is natural to call our problem of determining  $M(m, r)$  the 2-density Turán problem. Note that since  $m_2(H) \geq d_2(H)$  the proposition also holds if we replace  $M$  with  $\min d_2(H)$ . This essentially recovers the argument of Linial and Rabinovich [30]. However, it turns out one can in many cases do much better by using the actual 2-density.

### 3.1 The 2-density Turán Problem

In this subsection, we show our results concerning the 2-density Turán problem of determining  $M(m, r)$  which together with Proposition 1.4 give upper bounds in the local to global independence number problem mentioned in the introduction.

#### 3.1.1 Triangle-Free Case

Here we show our bounds for the case  $k = 3$ . This means that  $m$  and  $r$  satisfy  $2r - 1 \leq m \leq 3r - 3$  and as expected the behaviour will be very different at the beginning and end of the range. Our first observation determines  $M(2r - 1, r)$ .

**Proposition 3.1** *Let  $r \geq 2$ . Then  $M(2r - 1, r) = m_2(C_{2r-1}) = 1 + \frac{1}{2r-3}$ .*

**Proof** Since  $C_{2r-1}$  is a  $2r - 1$  vertex graph with no independent set of size  $r$  we obtain  $M(2r - 1, r) \leq m_2(C_{2r-1})$ . For the lower bound let  $G$  be a graph on  $2r - 1$  vertices with  $\alpha(G) \leq r - 1$ , our goal is to show  $m_2(G) \geq m_2(C_{2r-1})$ . If  $G$  contains a cycle of length  $\ell$ , then  $m_2(G) \geq m_2(C_\ell) \geq m_2(C_{2r-1})$  where in the last inequality we used  $\ell \leq 2r - 1$  since  $G$  has only  $2r - 1$  vertices. If  $G$  contains no cycles it is a forest so in particular it is bipartite. One part of the bipartition must have at least  $r$  vertices giving us an independent set of size at least  $r$ , which is a contradiction.  $\square$

Turning to the other end of the range we show.

**Theorem 3.2** For  $r \geq 2$  we have  $M(3r - 4, r) \geq \frac{5}{3} - \frac{1}{r-2}$ .

**Proof** Let  $G$  be a graph on  $m = 3r - 4$  vertices with  $\alpha(G) \leq r - 1$ . If  $G$  contains a triangle, then  $m_2(G) \geq 2$  and we are done. If  $G$  contains a subgraph  $G'$  on  $m'$  vertices with minimum degree at least 4, then  $m_2(G) \geq d_2(G') \geq \frac{2m'-1}{m'-2} > 2$  so again we are done. In particular, we may assume that  $G$  is 3-degenerate. These conditions allow us to apply a modification of a result of Jones [21] (see Appendix A for more details about the modification) which tells us that  $e(G) \geq 6m - 13(r - 1) - 1 = 5r - 12$ . This implies  $m_2(G) \geq d_2(G) \geq \frac{5r-13}{3r-6} = \frac{5}{3} - \frac{1}{r-2}$  as claimed.  $\square$

This is close to the best possible, for example, the chain graph  $H_r$  (see [21] for more details) has  $3r - 4$  vertices, no independent set of size  $r$  and  $m_2(H_r) = \frac{5}{3} - \frac{1}{9} \cdot \frac{1}{r-2}$ . We believe that as in the problem of [21], these graphs should be optimal, it is not hard to verify that this is indeed the case for the first few values of  $r$  and one can improve our result above by repeating more carefully the stability type argument from [21] for our graphs.

In the above result, we did not look at the very end of the range for  $k = 3$ , namely  $m = 3r - 3$ . The reason is that it seems to behave differently. Of course,  $M(3r - 3, r) \geq M(3r - 4, r)$  so the same bound as above applies, however, it seems possible that a stronger bound is the actual truth, it is even possible that the answer jumps to  $M(3r - 3, r) \geq 2$ .

### 3.1.2 Independence Number Two

In this subsection, we solve the 2-density Turán problem for graphs with no independent sets of size 3. The behaviour depends on the parity of  $m$ , we begin with the easier case when  $m$  is even.

**Lemma 3.3** For any  $k \geq 2$  we have  $M(2k, 3) \geq (k + 1)/2$ .

**Proof** Let  $G$  be a graph on  $2k$  vertices with  $\alpha(G) \leq 2$ . This condition implies that for any vertex  $v$  of  $G$  the set of vertices not adjacent to  $v$  must span a clique since otherwise, the missing edge together with  $v$  makes an independent set in  $G$  of size 3. On the other hand, if we can find  $K_k \subseteq G$ , then  $m_2(G) \geq m_2(K_k) = \frac{k+1}{2}$  and we are done. So we may assume  $G$  is  $K_k$ -free. Combining these two observations implies every vertex has at most  $k - 1$  non-neighbours and in particular  $\delta(G) \geq 2k - 1 - (k - 1) = k$ . This in turn implies  $m_2(G) \geq \frac{e(G)-1}{|G|-2} \geq \frac{k^2-1}{2k-2} = (k + 1)/2$  completing the proof.  $\square$

We now turn to the more involved case of odd  $m = 2k - 1$ . The increase in difficulty is partially due to the fact that the answer becomes very close (but not equal) to  $m_2(K_k)$  which we have seen above is the answer for graphs with one more vertex. So the bound we need to show is much stronger in the odd case. We begin with the following lemma which is at the heart of our argument. We state it for the complement of our actual graphs for convenience.

**Lemma 3.4** *Let  $k \geq 5$  and  $1 \leq t < \sqrt{(k - 1)/2}$ . Let  $G$  be a triangle-free graph on  $2k - 1$  vertices with the property that any  $k$  of its vertices span at least  $t + 1$  edges. Then  $e(G) \leq (k - 1)^2 - t^2 + 1$ .*

**Proof** The following easy claim will be used at various points in the proof. It also provides an illustration for the flavour of the more involved arguments we will be using later.

**Claim** *If there are 2 vertex disjoint independent sets of order  $k - 1$ , then  $e(G) \leq (k - 1)^2 - t^2 + 1$ .*

**Proof** Let  $v$  be the (only) vertex not belonging to either of the independent sets, which we call  $L$  and  $R$ . Let  $i$  denote the number of neighbours of  $v$  in  $L$  and  $j$  in  $R$ . Observe first that  $v \cup L$  and  $v \cup R$  are both sets of  $k$  vertices so need to span at least  $t + 1$  edges, by our main assumption on  $G$ . Since all edges in these sets are incident to  $v$  ( $L$  and  $R$  are both independent sets) we conclude that  $i, j \geq t + 1$ . Note further that, since  $G$  is triangle-free, there can be no edges between neighbours of  $v$ , which means that there can be at most  $(k - 1)^2 - ij$  edges in  $L \cup R = G \setminus v$ . Adding the  $i + j$  edges incident to  $v$  we obtain  $e(G) \leq i + j + (k - 1)^2 - ij = (k - 1)^2 + 1 - (i - 1)(j - 1) \leq (k - 1)^2 + 1 - t^2$ . □

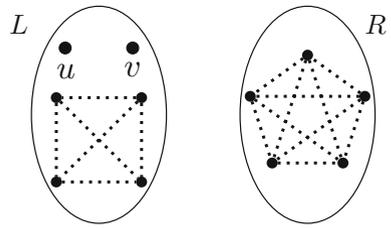
We now proceed to obtain some information on the structure of  $G$ . Observe first that by our assumption on  $t$  we have  $(k - 1)^2 - t^2 + 1 > (k - 1)^2 - (k - 1)/2 + 1 = k^2 - 5k/2 + 5/2$  so if we can show  $e(G) \leq k^2 - 5k/2 + 5/2$  we are done. So let us assume  $e(G) \geq k^2 - 5k/2 + 3$ , which will suffice to give us some preliminary information about  $G$ .

Since  $G$  is triangle-free, neighbours of any vertex span an independent set. By our main assumption on  $G$  there can be no independent set of order  $k$  so  $\Delta(G) \leq k - 1$ . On the other hand, we have  $\Delta(G) \geq 2e(G)/(2k - 1) \geq (2k^2 - 5k + 6)/(2k - 1) > k - 2$ , so  $\Delta(G) = k - 1$ . In particular, there exists a vertex with  $k - 1$  neighbours, which means that there is an independent set  $R$  of size  $k - 1$  in  $G$ . If every vertex in  $R$  has degree at most  $k - 3$ , then the sum of degrees in  $G$  is at most  $k(k - 1) + (k - 1)(k - 3) = 2k^2 - 5k + 3 < 2e(G)$ . So there is a vertex in  $R$  with degree at least  $k - 2$  and in particular, there exists an independent set of size at least  $k - 2$  disjoint from  $R$ . In other words,  $L := G \setminus R$  contains an independent set of size  $k - 2$  and two remaining vertices, say  $v$  and  $u$  (see Fig. 4 for the illustration of the current state). Let us w.l.o.g. assume that  $v$  has at most as many neighbours in  $L$  as  $u$ .

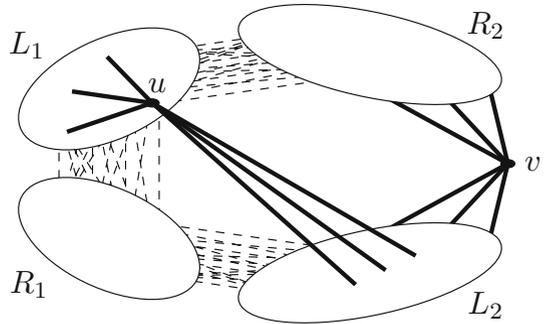
This almost gives us the situation in the claim above. In particular, by the claim, we may assume that both  $u$  and  $v$  have at least one neighbour in  $L \setminus \{u, v\}$ .

We now proceed to obtain more detailed information on how  $G$  should look like. Denote by  $L_2 := N(v) \cap L, R_2 := N(v) \cap R, L_1 = L \setminus (L_2 \cup \{v\})$  and  $R_1 := R \setminus R_2$ .

**Fig. 4** Initial structure for  $k = 6$ , dotted lines depict missing edges



**Fig. 5** Blow-up of  $C_5$  structure, dashed lines denote only possible locations of edges, there can be no edges between non-adjacent parts or inside parts, apart from those belonging to  $S$



Note that all edges within  $L_1 \cup L_2$  must touch  $u$  so  $L_1 \cup L_2$  induces a star  $S$  with a centre at  $u$ , say of size  $s$ . Note further that no edges in  $R_2 \cup L_2 = N(v)$  can exist as  $G$  is triangle-free. Putting these observations together we conclude that  $G$  without edges of  $S$  is a subgraph of a blow-up of  $C_5$  with parts  $\{v\}, R_2, L_1, R_1$  and  $L_2$  in order (see Fig. 5 for an illustration in the case when  $u \in L_1$ ).

This means that  $v$  contributes  $|L_2| + |R_2|$  edges while the remaining edges all come between  $L_1 \cup L_2$  and  $R_1 \cup R_2$  and  $S$ . Since  $|R_1 \cup R_2| = |L_1 \cup L_2| = k - 1$  and there can be no edges between  $R_2$  and  $L_2$  there are  $(k - 1)^2 - |L_2||R_2| - X$  edges between  $L_1 \cup L_2$  and  $R_1 \cup R_2$  where  $X$  denotes the number of non-edges between  $R_2$  and  $L_1, L_1$  and  $R_1$ , and  $R_1$  and  $L_2$ . In total we have  $e(G) = |L_2| + |R_2| + (k - 1)^2 - |L_2||R_2| - X + s$ .

Let us denote by  $i = |R_2|$  and  $j = |L_2|$  so  $k - 1 - j = |L_1|$ ;  $k - 1 - i = |R_1|$  and  $e(G) = i + j + (k - 1)^2 - ij - X + s = (k - 1)^2 + 1 - (i - 1)(j - 1) - X + s$ . Since by our main assumption on  $G$  there needs to be at least  $t + 1$  edges among the  $k$  vertices  $v \cup R_1 \cup R_2$  and we know there are exactly  $|R_2| = i$  edges in this set ( $R_1 \cup R_2$  is an independent set) we conclude that  $i \geq t + 1$ . Similarly, we know  $s + j \geq t + 1$  as otherwise  $v \cup L_1 \cup L_2$  make a  $k$ -vertex subset with  $|L_2| + s \leq t$  edges in total. Note that since we observed by the claim that both  $v$  and  $u$  need to have a neighbour inside  $L \setminus \{u, v\}$  we must have  $j, s \geq 1$ , while by our choice of  $v$  as having fewer neighbours in  $L$  than  $u$  we have  $j \leq s$ .

We distinguish two cases depending on whether  $v \sim u$  or not (i.e. whether  $u \in L_1$  or  $u \in L_2$ ). Let us deal with the case  $u \in L_2$  first. There can be no edges within  $L_2 = N(v)$  so all leaves of  $S$  must be in  $L_1$ .  $u$  must have at least  $t + 1$  neighbours within  $R_1 \cup R_2$  (otherwise  $u \cup R_1 \cup R_2$  are  $k$  vertices with at most  $t$  edges) so there must be at least  $s(t + 1)$  edges missing between  $R_1 \cup R_2$  and  $L_1$ , i.e.  $X \geq s(t + 1)$ . In particular,  $e(G) \leq (k - 1)^2 + 1 - (i - 1)(j - 1) + s - s(t + 1) \leq (k - 1)^2 + 1 - t(j - 1) - ts \leq$

$(k - 1)^2 + 1 - t^2$ , (where we used  $j \geq 1$  and  $i \geq t + 1$  in the second inequality and  $j - 1 + s \geq t$  in the third), as desired.

In the remaining case  $u \in L_1$ . Let's say  $s_1$  leaves of  $S$  are in  $L_1$  and  $s_2$  in  $L_2$ . Let  $x \geq t + 1$  be the number of neighbours of  $u$  in  $R_1 \cup R_2$  we know as before there must be  $xs_1 + (k - 1 - x)$  non-edges between  $L_1$  and  $R_1 \cup R_2$  and at least  $(x - |R_2|)s_2 = (x - i)s_2$  non-edges between  $R_1$  and  $L_2$ . In total we have  $X \geq xs_1 + (k - 1 - x) + \max(x - i, 0) \cdot s_2 = k - 2 + s + (x - 1)(s - 1) - \min(x, i) \cdot s_2$ . If we denote by  $m := \min(x, i)$  we get

$$\begin{aligned} e(G) &\leq (k - 1)^2 + 1 - (k - 2) - (i - 1)(j - 1) - (x - 1)(s - 1) + m \cdot s_2 \\ &\leq (k - 1)^2 + 1 - (k - 2) - (m - 1)(s + j - 2) + m \cdot s_2 \\ &\leq (k - 1)^2 + 1 - (k - 2) - (m - 1)(s + s_2 - 2) + m \cdot s_2 \\ &= (k - 1)^2 + 1 - (k - 4) - (m - 2)(s - 2) - s_1 \end{aligned}$$

Where we used  $i, x \geq m$  in the second inequality and  $j \geq s_2$  (since  $s_2$  leaves live inside  $L_2$  of size  $j$ ) in the third. The term  $(m - 2)(s - 2) + s_1$  is non-negative provided  $s \geq 2$ , since  $m = \min(x, i) \geq t + 1 \geq 2$ . So if  $s \geq 2$ , we have  $e \leq (k - 1)^2 + 1 - (k - 4) \leq (k - 1)^2 + 1 - t^2$ , where we used  $t < \sqrt{(k - 1)/2} \implies t^2 \leq k - 4$ , which holds for  $k \geq 5$  (using integrality of  $t$  for  $k = 5, 6$ ). If  $s = 1$  we must also have  $j = 1$  (since  $j \leq s$  and  $j \geq 1$ ) and in turn  $j + s \geq t + 1$  implies  $t = 1$ . If  $s_2 = 0$  the first inequality above (and  $k \geq 3$ ) gives  $e(G) \leq (k - 1)^2 = (k - 1)^2 - t + 1$  and we are done. If  $s_2 = |L_2| = 1$ , then  $s_1 = 0$  and removing the single vertex in  $L_2$  removes all neighbours of  $v, u$  in  $L$  and gives us again the situation from the claim.  $\square$

We are now ready to deduce our bound on  $M(2k - 1, 3)$ .

**Theorem 3.5** *Let  $k \geq 4$ , then  $M(2k - 1, 3) \geq \frac{k+1}{2} - \max_{1 \leq t \leq k-2} \min\left(\frac{t}{k-2}, \frac{(k+1)/2 - (t-1)^2}{2k-3}\right)$ .*

**Proof** Let  $G$  be a graph with  $2k - 1$  vertices and  $\alpha(G) \leq 2$ . If  $k = 4$  the desired bound evaluates to 2. Since in this case  $G$  does not satisfy the 7-local 3-independence property by Lemma 3.6 it contains either a  $K_4$  or an  $H_7$  as a subgraph. Since  $m_2(K_4) = 5/2 > 2$  and  $m_2(H_7) \geq \frac{e(H_7)-1}{|H_7|-2} = 2$  (since  $e(H_7) = 11$  and  $|H_7| = 7$ ) we deduce that in either case  $m_2(G) \geq 2$  as desired. Let us now assume  $k \geq 5$ .

Let  $m(t) = \min\left(\frac{t}{k-2}, \frac{(k+1)/2 - (t-1)^2}{2k-3}\right)$ . Observe that if we choose  $t = \sqrt{(k - 1)/2}$ , we obtain  $\frac{(k+1)/2 - (t-1)^2}{2k-3} = \frac{t^2 + 1 - (t-1)^2}{2k-3} = \frac{2t}{2k-3} < \frac{t}{k-2}$ . Let  $a$  be the maximiser of  $m(t)$ .

If  $m(a) = \frac{a}{k-2}$ , then by the above observation we must have  $a < \sqrt{(k - 1)/2}$  (since the first term of the minimum is increasing and the second is decreasing in  $t$ ). We may assume that among any  $k$  vertices there are at least  $a + 1$  missing edges, as otherwise the induced subgraph on these  $k$  vertices implies  $m_2(G) \geq (k + 1)/2 - a/(k - 2) = (k + 1)/2 - m(a)$  and we are done. This allows us to apply Lemma 3.4 to the complement of  $G$  to deduce  $G$  must have at least  $\binom{2k-1}{2} - ((k - 1)^2 + 1 - a^2) = k(k - 1) + a^2 - 1$  edges. This in turn implies  $m_2(G) \geq \frac{k(k-1)+a^2-2}{2k-3} = (k + 1)/2 - \frac{(k+1)/2 - a^2}{2k-3} \geq (k + 1)/2 - m(a)$ , where the last inequality follows since otherwise  $m(a) < \frac{(k+1)/2 - a^2}{2k-3}$

which implies  $m(a) < m(a + 1)$  (since also  $\frac{a+1}{k-2} > \frac{a}{k-2} = m(a)$ ), which contradicts maximality of  $m(a)$  (note that  $a + 1 < \sqrt{(k - 1)/2} + 1 \leq k - 2$  for  $k \geq 5$ ).

If on the other hand  $m(a) = \frac{(k+1)/2-(a-1)^2}{2k-3}$  we may assume  $a \geq 2$  (since  $\frac{1}{k-2} < \frac{(k+1)/2}{2k-3}$  for  $k \geq 5$ ). So we must have  $m(a) \geq \frac{a-1}{k-2}$  (otherwise since the second term in the definition of  $m(t)$  is decreasing in  $t$  we conclude  $m(a - 1) > m(a)$  and get a contradiction). As in the previous case this means that any  $k$  vertices must miss at least  $a$  edges, or we are done. We also know that  $m(a - 1) = (a - 1)/(k - 2)$  (as otherwise again  $m(a - 1) > m(a)$ ) so again as in the previous case by our initial observation we must have  $a - 1 < \sqrt{(k - 1)/2}$  and we may apply Lemma 3.4 with  $t = a - 1$  to complement of  $G$  to obtain  $e(G) \geq k(k - 1) + (a - 1)^2 - 1$ . This implies

$$m_2(G) \geq \frac{k(k - 1) + (a - 1)^2 - 2}{2k - 3} = \frac{k + 1}{2} - \frac{(k + 1)/2 - (a - 1)^2}{2k - 3} = \frac{k + 1}{2} - m(a), \quad (4)$$

as desired. □

We now prove the lemma which we used for the  $k = 4$  case in the above theorem and mentioned in Sect. 2.

**Lemma 3.6** *Any  $K_4$  and  $H_7$ -free graph  $G$  satisfies  $\alpha_7(G) \geq 3$ .*

**Proof** It is enough to show that any 7 vertex graph  $G$  which is  $K_4$ -free and has no  $I_3$  (independent set of size 3) must contain  $H_7$ . First observe that  $\delta(G) \geq 3$  as otherwise there is a vertex with 4 non-neighbours who must span a  $K_4$ , in order to avoid  $I_3$ . Note also that  $\Delta(G) \leq 5$  as if a vertex  $v$  had degree 6, then by  $R(3, 3) = 6$  in its neighbourhood we find an  $I_3$  or a  $K_3$  which together with  $v$  makes a  $K_4$ . Since  $G$  has odd size it must contain a vertex  $v$  of degree exactly 4.

Our goal is to find two vertex disjoint triangles in  $G$ . If some vertex  $v$  has degree 3, then its non-neighbours span a triangle and since its neighbours don't span an  $I_3$  the edge among them together with  $v$  give us our second triangle. If all vertices have degree at least 4, then by pigeonhole principle any two adjacent vertices lie in a triangle. If we take  $v$  as our guaranteed vertex of degree 4, let  $u, w$  be its non-neighbours. Then  $u \sim w$  and  $u$  and  $w$  lie in some triangle. Removing it leaves us with  $v$  and 3 of its neighbours, so we again find a second triangle.

So we can always find a pair of vertex disjoint triangles  $xyz$  and  $abc$ . Let  $v$  be the remaining vertex of  $G$ .  $v$  can send at most 2 edges towards each of the triangles. If it sends exactly 2 to both, then the non-neighbours of  $v$  must be adjacent and we found our  $H_7$ . If it sends 2 to  $xyz$ , say  $v \sim x, y$  of them but only one to  $abc$  say  $v \sim a$ , then replacing  $v$  with  $z$  we find disjoint triangles  $xyv$  and  $abc$  such that  $z$  sends 2 edges towards each ( $v \approx z, b, c$  implies  $z \sim b, c$ ) so we are back in the first case and are done. □

Combining Lemma 3.3 and Theorem 3.5 with Proposition 1.4 we obtain Theorem 1.6. Both above results are tight. Since we only needed the above lower bounds for Theorem 1.6 we will only describe our tightness examples here and postpone (the somewhat tedious) computation of their 2-density to Appendix C.

If  $m = 2k$ , then our example is simply a vertex disjoint union of 2 cliques on  $k$  vertices. This graph has no independent set of size 3 and it is not hard to see

that its 2-density is equal to  $m_2(K_k) = \frac{k+1}{2}$  (see Lemma C.2). If  $m = 2k - 1$  the answer is more complicated since it needs to match the somewhat messy bound of Theorem 3.5. The examples however still arise naturally from looking at the proof and will be blow-ups of  $C_5$  with cliques placed into parts which we choose to have sizes  $1, a, k - 1 - a, k - 1 - a, a$  in order around the cycle, where  $a$  is the optimal choice of  $t$  in Theorem 3.5. The complement of any such graph is an actual blow-up of  $C_5$  so is triangle-free and for the computation of its 2-density see Lemma C.3.

### 3.1.3 General Turán 2-density Problem

In this section, we show our general bounds on  $M(m, r)$ . Combining the following proposition with Proposition 1.4 we obtain Theorem 1.7.

**Proposition 3.7** *Let  $k = \lceil m/(r - 1) \rceil$ . Provided  $m$  is sufficiently larger than  $r$  we have  $M(m, r) \geq \frac{k+1}{2} - \frac{c_r}{\sqrt{k}}$ , where  $c_r > 0$  is a constant depending only on  $r$ .*

**Proof** Let  $G$  be an  $m$ -vertex graph with  $\alpha(G) \leq r - 1$ , our task is to show  $m_2(G) \geq \frac{k+1}{2} - \frac{c_r}{\sqrt{k}}$ , where  $k = \lceil \frac{m}{r-1} \rceil$ , under the assumption that  $m$  is large.  $\alpha(G) \leq r - 1$  implies that  $G$ 's complement  $\bar{G}$  is  $K_r$ -free. Let  $t = t_{r-1}(m) - e(\bar{G})$ , where  $t_{r-1}(m)$  denotes the Turán number for  $K_r$ -free graphs on  $m$  vertices. If  $t \geq \frac{3}{2}m$ , then we get

$$\begin{aligned} m_2(G) &\geq \frac{e(G) - 1}{|G| - 2} = \frac{\binom{m}{2} - t_{r-1}(m) + t}{m - 2} \geq \frac{\frac{m^2}{2(r-1)} + t - m/2}{m} \geq \frac{m}{2(r - 1)} + 1 \\ &\geq \frac{1}{2} \cdot (k - 1) + 1 = \frac{k + 1}{2}, \end{aligned}$$

where in the second inequality we used the standard bound  $t_{r-1}(m) \leq \left(1 - \frac{1}{r-1}\right) \frac{m^2}{2}$ .

So we are done unless  $t < \frac{3}{2}m$ . Since  $\bar{G}$  is  $K_r$ -free and has  $t_{r-1}(m) - t$  edges a stability theorem (see Theorem 1.3 in [6]) implies that  $\bar{G}$  can be made  $r - 1$ -partite by removing at most  $\frac{r t^{3/2}}{2m}$  edges (being crude and using that  $m$  is sufficiently larger than  $r$ ). Translating this to  $G$  we conclude  $G$  is a vertex disjoint union of  $r - 1$  cliques missing a few edges, in total at most  $\frac{r t^{3/2}}{2m} \leq r \sqrt{m}$  edges. At least one of these ‘‘cliques’’ needs to have size  $s \geq k$ . In particular, if we take a subset of  $k$  vertices of this ‘‘clique,’’ it still misses at most  $r \sqrt{m}$  edges. In particular, it has 2-density at least  $\frac{k+1}{2} - \frac{r\sqrt{m}}{k-2} \geq \frac{k+1}{2} - \frac{c_r}{\sqrt{k}}$ .

**Remark** The stability result we used above was also independently discovered in [33] (in an asymptotic form), we used the variant from [6] since it is explicit. Our problem seems to be closely related to this type of stability problems for Turán’s theorem. For example, the bipartite variant, which was precisely solved in [16], has the same form of optimal examples as we found for  $M(2k - 1, 3)$ . This was recently generalised to  $r$ -partite graphs in [23] which might be helpful for studying  $M(m, r)$  for larger  $r$ .

Note that for the special case of  $r = 3$  and  $m$  odd this result matches (up to a constant factor in front of the lower order term) our bound in Theorem 3.5 and is

hence almost best possible in this case by Lemma C.3. On the other hand, if  $m$  is even it is some way off. This seems to happen in general, we found examples (disjoint unions of our examples for the  $r = 3$  case) which show that the above bound is tight up to the constant factor in front of the lower order term provided  $m \pmod{r-1}$  is between 1 and  $(r-1)/2$ . This condition ensures that in the Turán  $K_r$ -free graph on  $m$  vertices there are more small parts (of size  $k-1$ ) which allows us to pair up small and big parts and place there a copy of our example from the  $r = 3$  case, we once again relegate the details to Appendix C. It seems that as  $m \pmod{r-1}$  approaches  $r$  stronger bounds should hold and ultimately if  $m \mid (r-1)$ , the lower order term disappears completely as it did in the  $r = 3$  case.

Theorem 1.7, while being close to best possible, unfortunately requires  $m$  to be somewhat large (compared to  $r$ ) which misses many interesting instances of the problem. The following result illustrates some of our ideas for obtaining results which hold for any choice of parameters. We restrict attention to the divisible case  $m = k(r-1)$  to keep the argument as simple as possible.

**Proposition 3.8** *Let  $r \geq 3$  and  $m = k(r-1)$ , then we have  $M(m, r) \geq \frac{k}{2} + \frac{k-1}{m-2}$ .*

**Proof** We will prove the result by induction on  $r$  while keeping  $k$  fixed. For the base case of  $r = 3$ , we have  $k = m/2$  and the statement matches precisely Lemma 3.3.

Let  $G$  be a graph on  $m$  vertices having  $\alpha(G) \leq r-1$ . If  $G$  has a vertex  $v$  of degree less than  $k$ , then removing  $v \cup N(v)$  from  $G$  we obtain a graph  $G'$  on at least  $m-k = k(r-2)$  vertices which has  $\alpha(G') \leq r-2$ , since  $v$  extends any independent set we can find in  $G'$ . This implies by the inductive assumption for  $r-1$  that  $m_2(G) \geq m_2(G') \geq M(m-k, r-1) \geq \frac{k}{2} + \frac{k-1}{m-k-2} \geq \frac{k}{2} + \frac{k-1}{m-2}$  with room to spare. Hence, we may assume  $\delta(G) \geq k$  which implies  $m_2(G) \geq d_2(G) \geq \frac{mk/2-1}{m-2} = \frac{k}{2} + \frac{k-1}{m-2}$ .  $\square$

The above proof clearly leaves quite some room for improvement. However, it (combined with  $M(m, r)$  being increasing in  $m$  to capture the non-divisible cases) already suffices to improve the bound of Linial and Rabinovich for all values of  $m$  and  $r$  with  $k \geq 4$ . It also suffices to obtain a significant improvement in the benchmark case  $m = 20, r = 5$  of  $M(20, 5) \geq 49/18$  over the previously best bound of  $43/18$  of Kostochka and Yancey [24]. We have more involved ideas which allow one to improve on the above bound quite substantially. In particular, we manage to resolve the benchmark case and show  $M(20, 5) = 3$ . Since that argument is somewhat more involved and its generalisations become even more complicated, while ultimately still falling short of the asymptotic result of Theorem 1.7, we relegate it to Appendix B.

## 4 Concluding Remarks and Open Problems

In this paper we study the local to global independence number problem, i.e. how big an independent set one finds in a graph with the property that any  $m$  vertices contain an independent set of size  $r$ . While many of our results break previous barriers on this problem, there is still room for improvement and we believe we have not fully exhausted the potential of our ideas.

In terms of lower bounds, we improve previously best bounds for about half of the possible choices of  $m$  and  $r$ . It would be interesting to obtain a similar improvement for the whole, or at least most of the range. Our argument here relied on improving the bounds for  $r = 3$  which is then generalised through Lemma 2.1. One can follow our approach for  $r \geq 4$  as well. Lemma 2.1 easily generalises so for example if one improves bounds say for  $m = 3k - 1$ ,  $r = 4$ , this leads to improvement for about  $2/3$  of the possible values in general. For  $r = 3$  our arguments relied on a Ramsey result for graphs  $H_{2k-1}$  which were certain blow-ups of  $C_5$ . We believe a similar story should happen for larger  $r$ , in the initial cases role of  $C_5$  could be taken by the chain graphs (see graphs  $H_k$  in [21]) and to obtain a general result for fixed  $r$  one should prove a Ramsey bound for appropriate blow-ups of these chain graphs. This should lead to an improvement for essentially all values of  $m$  and  $r$  except when  $r - 1 \mid m$  which seems more difficult. In fact, using a minor modification of Lemma 2.1 if one improves the bounds in such a “divisible” case, say  $m = 2k$  and  $r = 3$ , this immediately improves the bounds for any choice of  $m$  and  $r$  with  $r - 1 \nmid m$  (and most divisible cases as well). Here a good starting point seems to be the case  $m = 8$ ,  $r = 3$ .

**Question 4.1** *Does any graph with an independent set of size 3 among any 8 vertices have  $\alpha(G) \geq n^{1/3+\varepsilon}$ ?*

The reason we raise the  $(8, 3)$  case instead of  $(6, 3)$  is that the latter is easily seen to be essentially equivalent to the problem of how large independent sets we find in triangle-free graphs. Since the answer to this classical problem is known up to a constant factor [7, 17, 34] the same holds for our problem. This raises the possibility that the  $(8, 3)$  case is essentially equivalent to the same problem for  $K_4$ -free graphs which is open and believed hard. It turns out however that this is not the case since for example the square of  $C_8$  is an 8 vertex  $K_4$ -free graph with no independent set of size 3 and could play the role of our  $H_{2k-1}$ 's as the intermediate forbidden graph in this case. In fact, it is the only possible candidate, as can be seen by looking at optimal examples for  $R(4, 3)$  [31] which show that property  $\alpha_8 \geq 3$  is essentially equivalent to the graph being  $K_4$ -free and  $C_8^2$ -free.

In terms of improving our new bounds a good starting place are graphs in which every 7 vertices have an independent set of size 3. We showed such graphs must have  $\alpha(G) \geq n^{5/12-o(1)}$  proving a conjecture of Erdős and Hajnal. Here the natural limit for our methods is actually  $n^{3/7}$  and most of our argument works up to this point. It should be possible to push our methods at least beyond  $5/12$ . On the other hand, breaking  $3/7$  seems to require new ideas. The main question here is whether it is possible to reach  $1/2$ , namely whether the second conjecture of Erdős and Hajnal holds.

**Question 4.2** *Does any graph with an independent set of size 3 among any 7 vertices have  $\alpha(G) \geq n^{1/2-o(1)}$ ?*

Lemma 3.6 shows that this is in some sense equivalent to a Ramsey problem of our graph  $H_7$  vs an independent set, with the added benefit that we know the graph is  $K_4$ -free, which however seems to be a weaker condition than being  $H_7$ -free, so it is unclear if it is actually needed at all. The above bound for  $m = 7$ ,  $r = 3$  is stronger than our general bound which makes it likely that the general bound can be further improved.

Let us now turn to the upper bounds. Our bounds all arise from our results on the Turán 2-density problem and the main open problem is to solve this problem precisely for all choices of parameters.

**Question 4.3** *What is the minimum value of the 2-density of a graph on  $m$  vertices having no independent set of size  $r$ ?*

We defined the answer to be  $M(m, r)$  and determine it precisely for  $r = 3$ , for ends of the range with  $k = 3$  (for  $m = 2r - 1$  and up to lower order term for  $m = 3r - 4$ ), for certain small cases such as  $m = 20, r = 5$  and determine it up to  $O_r(1/\sqrt{m})$  in general. While the parameter  $k = \lceil \frac{m}{r-1} \rceil$  seems to control the rough behaviour of  $M(m, r)$ , in order to obtain precise results one needs to take into account the residue of  $m$  modulo  $r - 1$ . We have seen this in the  $r = 3$  case with the distinction between even and odd cases. This is also evident in the  $k = 3$  case from our results for the ends of this range. The behaviour for  $k = 3$  across the whole range also seems interesting and may be a good starting point for obtaining precise general results.

In general, we can show that for the first half of the non-zero residues the  $O_r(1/\sqrt{m})$  term is needed. It could be interesting to determine what happens for the remaining half of the residues and in particular when  $r - 1 \mid m$ . Based on our results for  $r = 3$  and  $m = 20, r = 5$ , it seems plausible that  $M(m, r) = m_2(K_k) = \frac{k+1}{2}$  for any  $r$  and  $m = k(r - 1)$ .

Finally, let us summarise the current state of the art in the following table, where we remind the reader that  $f(n, m, r)$  stands for the smallest possible size of  $\alpha(G)$  in an  $n$  vertex graph with  $\alpha_m(G) \geq r$ .

$m$	$2r - 1$	$[2r, 3r - 5]$	$3r - 4, 3r - 3$	$\left( (k - 1)(r - 1), (k - \frac{1}{2})(r - 1) \right)$	$\left( (k - \frac{1}{2})(r - 1), k(r - 1) \right)$
$f(n, m, r) \geq$	$\Omega\left(n^{1-\frac{1}{r}}\right)$	$\Omega\left(n^{1-1/\lfloor \frac{m-r+1}{m-2r+2} \rfloor}\right)$	$\Omega\left(n^{\frac{1}{k-3/2}}\right)$	$\Omega\left(n^{\frac{1}{k-1}}\right)$	
$f(n, m, r) \leq$	$O\left(n^{1-\frac{1}{2r-2}}\right)$	$n^{\frac{3}{5} + \frac{2}{5r-13} + o(1)}$	$n^{\frac{2+o(1)}{k+1-O_r(1/\sqrt{k})}}$		

Where all the asymptotics are in terms of  $n \rightarrow \infty$  apart from the  $O_r$  term where  $r$  is a constant and  $k \rightarrow \infty$ .

We note in addition that the value for  $r \leq m \leq 2r - 2$  is known precisely and is equal to  $n - o(n)$ . One can use our arguments to obtain improvements for various parts of the above regimes and special cases, but let us mention here specifically that we have slightly weaker upper bounds in the final regime which do not require  $k$  to be large compared to  $r$  (see Proposition 3.8).

**Acknowledgements** We would like to thank Alexandr Kostochka for helping us find [21] and Michael Krivelevich for useful comments. We would also like to thank the anonymous referees for their careful reading of the paper and numerous suggestions which improved the presentation of this paper. The first author would like to gratefully acknowledge the support of the Oswald Veblen Fund.

**Funding** Open access funding provided by Swiss Federal Institute of Technology Zurich Open access funding provided by Swiss Federal Institute of Technology Zurich.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A: Independence Number in Sparse Triangle-Free Graphs

**Theorem A.1** Any 3-degenerate, triangle-free graph  $G$  with  $m$  vertices and no independent set of size  $r$  has  $e(G) \geq 6m - 13r - 1$ .

**Proof** Our proof is by induction on  $m$ . We will actually prove a slightly stronger result. We will show that  $e(G) \geq 6m - 13r - 1 + I$ , where  $I$  is equal to 1 if  $\delta(G) < 3$  and equal to 0 otherwise.

For the base case we show the result for  $m \leq 5$  and any  $r$ . If  $m \leq 2$ , the result is immediate since the desired bound follows from  $e(G) \geq 0$ . If  $m \geq 3$ , since the graph is triangle-free we must have  $r \geq 2$ , so the bound again follows for  $m \leq 4$  immediately and if  $r \geq 3$  also for  $m = 5$ . In the only remaining case, if  $r = 2$ ,  $m = 5$ , we must have  $G = C_5$  and the bound again holds.

Now let us assume  $m \geq 6$  and that the result holds for any graph on at most  $m - 1$  vertices satisfying our conditions. If there is an isolated vertex in  $G$ , then by removing it we obtain a graph on  $m - 1$  vertices with independence number at most  $r - 1$  which implies  $e(G) \geq 6(m - 1) - 13(r - 1) - 1 \geq 6m - 13r + 6 \geq 6m - 13r - 1 + I$ . Similarly if there is a vertex of degree 1, removing it and its neighbour leaves us with a graph on  $m - 2$  vertices with independence number at most  $r - 1$  giving us the bound  $e(G) \geq 6(m - 2) - 13(r - 1) - 1 \geq 6m - 13r \geq 6m - 13r - 1 + I$ . So we may assume  $\delta(G) \geq 2$ .

If  $\delta(G) = 2$  let  $v$  be a vertex of degree 2, with neighbours  $u, w$ . If there are at most 2 vertices, other than  $v$ , adjacent to one of  $u$  or  $w$ , then removing them and  $v, u, w$  leaves us with a graph on at least  $m - 5$  vertices with independence number at most  $r - 2$ . This leftover graph has at least  $6(m - 5) - 13(r - 2) - 1 = 6m - 13r - 5$  edges. On the other hand  $G$  in addition has at least 5 edges touching these 5 removed vertices since minimum degree is 2 giving us  $e(G) \geq 6m - 13r \geq 6m - 13r + I - 1$  as desired. So there are at least 3 vertices (other than  $v$ ) adjacent to  $u$  or  $w$  and in particular  $v, u, w$  are incident to at least 5 edges. By removing  $v, u$  and  $w$  we obtain a graph on  $m - 3$  vertices with independence number at most  $r - 1$  which hence has at least  $6(m - 3) - 13(r - 1) - 1 = 6m - 13r - 6$  edges. Since  $v, u, w$  are incident to at least 5 edges we need to gain one more. Note first that  $G \setminus u, v, w$  must have minimum degree 3 or we gain one from its  $I$  term. Note also that either  $u$  or  $w$  must have degree 2 or  $v, u, w$  touch at least 6 edges and we gain. Say  $u$  is of degree 2, and  $w'$  is its neighbour other than  $w$ .  $w'$  must have degree at least 3 in  $G \setminus u, v, w$  so at

least 4 in  $G$ . This means that  $u, v, w'$  touch at least 6 edges and repeating the above argument with  $u$  in place of  $v$  we are done.

Final case is if  $\delta(G) = 3$ . We may assume  $G$  is connected, as otherwise we may apply induction on each of the components and are done. If  $G$  is 3-regular, then the number of edges is  $e(G) = 3m/2$  and a result of Staton (see Theorem 6 in [35]) on graphs with maximum degree 3 implies  $r \geq \frac{5}{14}m$  this implies  $e(G) = 3m/2 \geq 6m - 13r$  as desired. Hence, we may assume there is a vertex of degree at least 4 and in particular, since  $G$  is 3-degenerate that there exists a vertex  $v$  of degree 3 adjacent to a vertex  $u$  of degree at least 4. Let  $w, q$  be the remaining neighbours of  $v$ . Since  $\delta(G) = 3$  and  $d(u) = 4$  we know there are at least 10 edges touching  $v, u, w$  or  $q$ . Removing these 4 vertices we get a graph on  $m-4$  vertices with independence number at most  $r-1$  so by induction it has at least  $6(m-4) - 13(r-1) + I' - 1 = 6m - 13r - 1 - 11 + I'$  edges, where  $I' = 1$  if the remainder graph has minimum degree at most 2. This means that we are done unless there are exactly 10 edges touching  $v, u, w, q$  and the remainder graph has minimum degree at least 3.  $w$  is a vertex of degree 3 with two neighbours in the remainder graph, so each having degree at least 3 there and at least 4 in  $G$  (since they are adjacent to  $w$ ). This means that  $w$  and its neighbours touch at least 11 edges so repeating the argument as above with  $w$  in place of  $v$  we obtain the desired bound.  $\square$

## Appendix B: The $M(20,5)$ Case

In order to show  $M(20, 5) \geq 3$ , we will need a few intermediate results. Let us define  $e(m, r)$  to be the minimum possible number of edges in an  $m$ -vertex graph with independence number at most  $r - 1$ , provided it has 2-density less than  $\frac{k+1}{2}$  (where  $k = \lceil m/(r - 1) \rceil$ ). It can be thought of as the variant of determining  $M(m, r)$  where we care about the final number of edges instead of the 2-density, but we impose a restriction of not having too dense parts. For determining  $M(20, 5)$  we will need bounds on  $e(14, 4)$  and  $M(15, 4)$  for which in turn we will need  $e(9, 3)$ .

**Lemma B.1**  $e(9, 3) \geq 19$ .

**Proof** Let  $G$  be a graph with 9-vertices with  $\alpha(G) \leq 2$  and  $m_2(G) < 3$ . If  $G$  has a vertex  $v$  of degree at most 3, then  $v$  has at least 5 non-neighbours which must span a clique (since  $\alpha(G) \leq 2$ ) which implies  $m_2(G) \geq m_2(K_5) = 3$ , a contradiction. If  $\delta(G) \geq 5$ , then  $e(G) \geq 9 \cdot 5/2 > 19$ , and we are done. So  $\delta(G) = 4$  and there exists a vertex  $v$  with degree exactly 4. Let  $L = v \cup N(v)$  and  $R = G \setminus L$ , so  $|L| = 5$ ,  $|R| = 4$ . Since  $|L| = 5$  there must be a missing edge in  $L$  (or we find a  $K_5$  and have  $m_2(G) \geq 3$ ). Observe that vertices making a missing edge in  $L$  can not both be non-adjacent to the same vertex in  $R$  (or  $\alpha(G) \geq 3$ ) meaning they need to send at least 4 edges towards  $R$ . We obtain that there must be at least  $4 + 3 \cdot 4 - 3 = 13$  edges touching  $L$ , since there are 4 cross edges touching the missing edge, the remaining 3 vertices inside  $L$  each have degree at least  $\delta(G) = 4$  and we double counted only the edges between these 3 vertices, so at most 3. Now since  $R$  consists of non-neighbours of  $v$  it spans a  $K_4$  so there are 6 edges within  $R$  and in total we have the claimed  $13 + 6 = 19$  edges.

**Lemma B.2**  $e(14, 4) \geq 33$ .

*Proof* Let  $G$  be a graph with 14-vertices with  $\alpha(G) \leq 3$  and  $m_2(G) < 3$ . If  $G$  has a vertex  $v$  of degree at most 3, then  $v$  has at least 10 non-neighbours which contain no independent set of size 2 so must have 2-density at least 3 by Lemma 3.3. If  $\delta(G) \geq 5$ , then  $e(G) \geq 14 \cdot 5/2 > 33$  as desired. So  $\delta(G) = 4$  and there exists a vertex  $v$  with degree exactly 4. Let again  $L = v \cup N(v)$  and  $R$  be the rest of the graph. Note now that a missing edge in  $L$  must touch at least 5 edges going to  $R$ , as otherwise there are  $9 - 4 \geq 5$  common non-neighbours of the missing edge and they must span a clique. Similarly as in Lemma B.1 this means there needs to be 14 edges touching  $L$ . On the other hand  $G[R]$  is a 9-vertex graph with  $\alpha(G[R]) \leq 3$  (since  $v$  is a non neighbour to anyone in  $R$ ) and it has  $m_2(G[r]) \leq m_2(G) < 3$  so Lemma B.1 applies implying there are at least 19 edges inside  $G[R]$  and giving us our claimed total.  $\square$

Backtracking along the above proofs it is not hard to show that they are both optimal and even deduce a lot about the structure of the optimal examples. We are now ready to prove our final intermediate result which determines  $M(15, 4)$ .

**Lemma B.3**  $M(15, 4) = 3$ .

*Proof* Let  $G$  be a graph with 15-vertices with  $\alpha(G) \leq 3$  and  $m_2(G) < 3$ . If there is a vertex of degree at most 4 removing it and its neighbourhood, similarly as above reduces to the case of  $M(10, 3)$  which by Lemma 3.3 is equal to 3. So  $\delta(G) \geq 5$ . If  $\delta(G) \geq 6$ , then  $e(G) \geq 45$  and  $d_2(G) \geq \frac{45-1}{15-2} > 3$ . So let  $v$  be a vertex with degree exactly 5. Let  $L = v \cup N(v)$  and  $R$  be the rest of the graph.  $G[R]$  is a 9-vertex graph with  $\alpha(G[R]) \leq 2$  and it has  $m_2(G[r]) \leq m_2(G) < 3$ , so  $R$  spans at least 19 edges by Lemma B.1. We claim there needs to be at least 21 edges touching  $L$ . If there were a triangle of missing edges inside  $L$  its vertices need to send at least 9 edges across (or we get  $\alpha(G) \geq 4$ ) so our usual calculation tells us there are  $9 + 3 \cdot 5 - 3 = 21$  edges touching  $L$  as desired. Note also that missing edges inside  $L$  can not span a star (or we can remove its centre and be left with a  $K_5$  inside  $L$  so there need to exist 2 disjoint missing edges. Each missing edge sends at least 5 edges across (or we find a  $K_5$  in their common non-neighbourhood) and since we are assuming there is no missing triangle inside  $L$  there needs to be 2 actual edges inside  $L$  between our missing pair. Putting these together we get  $2 \cdot 5 + 2 + 2 \cdot 5 - 1 = 21$  edges touching  $L$  as desired. So there are at least  $21 + 19 = 40$  edges in  $G$  and  $m_2(G) \geq 3$  as desired.  $\square$

**Theorem B.4**  $M(20, 5) = 3$ .

*Proof* Let  $G$  be a graph with 20-vertices with  $\alpha(G) \leq 4$  and  $m_2(G) < 3$ . If there is a vertex of degree at most 4 removing it and its neighbourhood, reduces to the case of  $M(15, 4)$  solved in Lemma B.3. So  $\delta(G) \geq 5$ . If there are more than 8 vertices of degree greater than 5, then  $e(G) \geq 55$  and  $d_2(G) \geq \frac{55-1}{20-2} = 3$ . So there are at least 12 vertices of degree 5. Let  $v$  be such a vertex and Let  $L = v \cup N(v)$  and  $R$  be the rest of the graph.  $G[R]$  is a 14-vertex graph with  $\alpha(G[R]) \leq 3$  and it has  $m_2(G[r]) \leq m_2(G) < 3$ , so  $R$  spans at least 33 edges by Lemma B.2. If we find 22 edges touching  $L$  we obtain  $e(G) \geq 55$  and are done.

Note that any missing triangle in  $L$  sends at least 10 edges across (or the  $\geq 5$  common non-neighbours need to make a clique). So as in the previous lemma we obtain at least 22 edges touching  $L$  as desired. Once again the missing edges can not span a star or we find a  $K_5$  inside  $L$  so again we need to be able to find a disjoint pair of missing edges in  $L$ . Since again any missing edge must send at least 5 edges to  $R$  (or we reduce to the case of  $M(10, 3) = 3$ ) and at least 2 actual edges must exist between the missing edges (or we find a missing triangle) so again we obtain  $2 \cdot 5 + 2 + 2 \cdot 5 - 1 = 21$  edges touching  $L$ .

This already implies  $M(20, 5) \geq 53/18$  but to obtain our best bound we need to work a little bit harder. In particular, if we do find only 21 edges touching  $L$  above we obtain a lot of structural information about  $L$ .

**Claim** *If there are 21 edges touching  $L$ , then  $G[L]$  induces either:*

- (a) *A  $K_3$  and a  $K_4$  intersecting in a single vertex and consisting entirely of vertices of degree 5 in  $G$  or*
- (b) *Two  $K_4$ 's intersecting in an edge, one  $K_4$  consists of vertices of degree 5 and the remaining 2 vertices are of degree 6.*

**Proof** We have already observed that there are no missing triangles in  $L$  and that we can find 2 disjoint missing edges, which gave us at least 21 edges touching  $L$ . In order for this bound to be tight there needs to be exactly two actual edges in between them and in order to avoid missing triangles this means that the 4 vertices making these two edges span a missing  $C_4$  with both cross edges present. Each of the 4 missing edges making this  $C_4$  must send exactly 5 edges across (or we again gain) and every vertex of this  $C_4$  has at most 3 neighbours in  $L$  so has at least 2 neighbours outside. The only way this can happen is if 2 diagonal (so adjacent in  $G$ ) vertices of  $L$  send 3 edges out (call them  $w_1, w_2$ ) and remaining 2 send 2 edges out (call them  $s_1, s_2$ ). By minimum degree in  $G$  being at least 5 we know that both  $s_i$  must be joined to both remaining vertices in  $L$  ( $v$  and call the final vertex  $u$ ). Since  $v$  is adjacent to everyone the only edges we don't have information about are  $uw_1, uw_2$ . If only  $uw_1$  is a missing edge, then  $u, w_1$  must send at least 5 edges across and since  $w_1$  sends 3  $u$  must send 2. But this would imply  $u$  is adjacent to  $v, w_2, s_1, s_2$  and these 2 outer vertices so have degree at least 6 and improve our bound. Therefore, there are only 2 options both  $uw_1$  and  $uw_2$  are edges, giving us the case (b) or neither are edges, giving us case (a). Note that we know  $u, v$  must be of degree 5 or we gain and since we know how many edges  $s_i$ 's and  $w_i$ 's send outside and how  $G[L]$  looks like we know degree's of everyone in  $G$ .  $\square$

Observe that  $v$  was an arbitrary vertex of degree 5 so the above must hold for any such vertex.

**Claim** *Case (b) happens for all vertices of degree 5.*

**Proof** Let us assume case (a) happens for vertex  $v$ . In other words  $v \cup N(v)$  induces a triangle  $v, u, w$  and a  $K_4 = v \cup U_v$ . Observe that  $v \in N(u)$  and  $v$  is not adjacent to 3 vertices in  $N(u)$  outside  $N(v)$  ( $u$  has degree 2 in  $v \cup N(v)$  and degree 5 in  $G$ ). This means  $u \cup N(u)$  must also fall under case (a) since in case (b) any vertex in  $G[u \cup N(u)]$  has at most 2 non-neighbours. So  $u \cup N(u)$  makes a triangle  $u, v, w$  plus

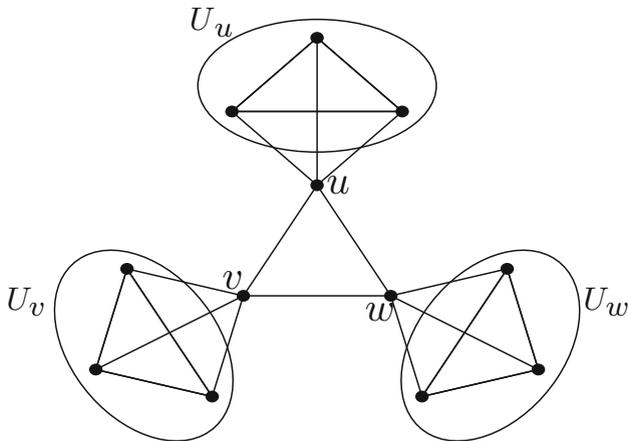


Fig. 6 Graph  $W$

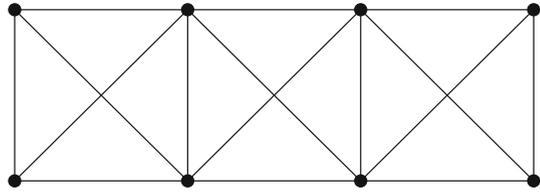
$K_4 = u \cup U_u$  (we know the triangles match since only the vertices of the triangle miss 3 edges in the neighbourhood, which means  $v$  is in the triangle and is only adjacent to  $u, w$ ). Analogously we obtain  $U_w$  with the same picture. So we find a graph  $W$  consisting of 3 vertex disjoint  $K_4$ 's each with a singled out vertex such that the singled out vertices make a triangle (see Fig. 6 for an illustration).

We know  $u, v, w$  have no further edges among vertices of  $W$  as they have degree 5 both in  $G$  and  $W$ . But we claim that also there are no edges between distinct  $U_v, U_u, U_w$ , meaning we find  $W$  as an induced subgraph. To see this if say  $x \in U_v, y \in U_u$  are adjacent, then they can send at most 1 edge outside  $W$  (they have degree 5 in  $G$  and 4 in  $W$ ). Since  $u$  and  $x$  are independent and there are at most 8 vertices adjacent to one of them ( $w, U_v, U_u$  and the potential neighbour of  $x$  outside  $W$ ) this means that removing  $u, N(u), x, N(x)$  removes at most 10 vertices and leaves us with a 10 vertex graph which has no independent set of size 3 (or together with  $u$  and  $x$  we get a size 5 one in  $G$ ) so this reduces to the  $M(10, 3)$  case and we are done.

Observe that every vertex of  $U_v, U_w, U_u$  has degree 3 in  $W$  and 5 in  $G$  so sends exactly 2 edges to  $G \setminus W$ . If there is  $x \in U_u$  and  $y \in U_w$  which have a common neighbour outside  $W$ , then there are at least 5 vertices outside of  $W$  which are not adjacent to either  $x$  or  $y$ . In particular, there is a missing edge among such vertices, which together with  $x, y$  and  $v$  makes an independent set of size 5. This means that  $N(U_u), N(U_w)$  and  $N(U_u)$  when restricted outside of  $W$  must be disjoint. In particular, since there are 8 remaining vertices there is one neighbourhood, say of  $U_v$  consisting only of 2 vertices. Since every vertex of  $U_v$  sends 2 edges outside  $W$  this means that  $U_v$  and these 2 vertices span a  $K_{3,2}$ . Finally, this means that for any vertex  $v'$  of  $U_v$  we have  $v' \cup N(v')$  induces a  $K_6$  minus a triangle, which falls under neither of our cases. This is a contradiction since  $v'$  has degree 5 in  $G$ .  $\square$

Let  $H$  denote the graph consisting of two vertex disjoint  $K_4$ 's with two pairs of vertices, one pair per part joined by a  $K_{2,2}$ . See Fig. 7 for an illustration.

Fig. 7 Graph  $H$



**Claim** Any vertex of degree 5 lies in an induced copy of  $H$  in  $G$  where vertices of degree 3 in  $H$  have degree 6 in  $G$ .

**Proof** Let  $v$  be a vertex of degree 5, since case (b) occurs we obtain two  $K_4$ 's sharing an edge  $uv$ . One of the  $K_4$ 's consists only of vertices of degree 5 so  $u, v$  and call its remaining 2 vertices  $w, s$ . So by looking at the neighbourhood of  $w$ , since case (b) must occur we deduce there is a  $K_4$  which intersects our two original cliques only in  $ws$  (since  $v, u$  would have too high degree if they were in the shared edge). This gives us a desired copy of  $H$  and it remains to be shown  $H$  is induced.

$u, v, s, w$  all have degree 5 in  $H$  and 5 in  $G$  so they have no additional edges. If we had an edge between  $x$  and  $y$ , both of which are of degree 6 and which satisfy  $x \sim w, y \sim v$ , then  $x$  has 4 edges in  $G[H]$  so sends at most 2 edges outside. Deleting these two edges and our copy of  $H$  we deleted at most 10 vertices among which we deleted  $x, N(x), v, N(v)$  and  $x$  and  $v$  are independent, so we again reduce to the  $M(10, 3)$  case and are done. □

**Claim** No two copies of  $H$  as in the previous claim can intersect.

**Proof** The central  $K_4$  in any such copy of  $H$  consists of vertices of degree 5 in  $G$  and in  $H$  so they have to be disjoint between copies of  $H$ .

If a vertex  $x$  of degree 6 is shared by two copies  $H_1, H_2$  of  $H$ , then  $x$  has either 6 distinct edges among  $H_1 \cup H_2$  or 5, but then  $|H_1 \cap H_2| \geq 2$ . Let us delete  $H_1, H_2$  and the potential external neighbour of  $x$ . In either case we deleted at most 15 vertices. Let us choose  $v_i \in H_i$  such that  $d(v_i) = 5$  and  $v_i \approx x$ . This means  $x, v_1, v_2$  make an independent set and we deleted them and their neighbourhoods in  $G$ . This means that among remaining 5 vertices there can be no missing edges as any such edge together with  $x, v_1, v_2$  would span a  $K_5$ . □

Finally this means that the copies of  $H$  we find should be vertex disjoint and since there are 12 vertices of degree 5 and each  $H$  contains 4 this means there should be at least 3 copies so at least 24 vertices in total giving us a contradiction and completing the proof. □

### Appendix C: Upper Bounds for the Turán 2-density Problem

In this section we show our upper bounds on  $M(m, r)$ . We begin with some basic observations which will prove useful in computing 2-densities. We say two vertices  $v$  and  $u$  are *equivalent* if there is a transposition swapping vertices  $v$  and  $u$  in the automorphism group of  $G$ .

**Lemma C.1** *Let  $H$  be the subgraph of  $G$  maximising  $d_2(G)$  which has as many vertices as possible. For any pair of equivalent vertices,  $H$  contains either both or none of them.*

**Proof** Let  $v$  and  $u$  be equivalent vertices and assume for the sake of contradiction that  $H$  contains  $v$  but not  $u$ . Let  $d$  denote the number of neighbours of  $v$  in  $H$ , in particular by equivalence  $u$  also has  $d$  neighbours in  $H \setminus \{v\}$ , so in particular at least  $d$  neighbours in  $H$ .

By maximality of  $H$  we must have  $\frac{e(H)-1}{|H|-2} \geq \frac{e(H)-d-1}{|H|-1-2} \implies d \geq \frac{e(H)-1}{|H|-2}$  or  $H \setminus \{v\}$  would have larger 2-density. Similarly, by comparing with  $H \cup \{u\}$  we must have  $\frac{e(H)-1}{|H|-2} > \frac{e(H)+d-1}{|H|+1-2} \implies d < \frac{e(H)-1}{|H|-2}$  giving us a contradiction.  $\square$

**Lemma C.2** *For any  $k \geq 2$  we have  $M(2k, 3) \leq \frac{k+1}{2}$ .*

**Proof** We take  $G$  to be a disjoint union of two  $K_k$ 's. It is immediate that  $\alpha(G) \leq 2$  so we only need to show  $m_2(G) = m_2(K_k) = \frac{k+1}{2}$ . To this end let us take the subgraph  $H$  of  $G$  as in Lemma C.1. Since all the vertices within a single clique are equivalent we deduce that  $H$  is either  $K_k$  or the whole graph. In the first case we are done immediately. In the second case we have  $m_2(G) = d_2(G) = \frac{k(k-1)-1}{2k-2} < \frac{k}{2} < \frac{k+1}{2}$  giving us a contradiction.  $\square$

**Remark** There are more interesting examples, for example one can find tight examples which are  $K_k$ -free. If  $k$  is even one can take  $k/2$ -th power of the cycle on  $2k$  vertices while for odd  $k$  one can find a different Cayley graph of the Cyclic group of order  $2k$ .

**Lemma C.3** *For any  $k \geq 4$  the bound in Theorem 3.5 is tight.*

**Proof** Let  $m(t) = \min\left(\frac{t}{k-2}, \frac{(k+1)/2-(t-1)^2}{2k-3}\right)$  and let  $a$  be the maximiser of this expression (over  $1 \leq t \leq k-2$ ), as in the proof of Theorem 3.5. Our goal is to find a graph  $G$  on  $2k-1$  vertices with  $\alpha(G) \leq 2$  which has  $m_2(G) \leq \frac{k+1}{2} - m(a)$ . We take  $G$  to be the blow-up of  $C_5$  with cliques placed inside of parts of size  $1, k-1-a, a, a, k-1-a$ , in order around the cycle. Since complement of  $G$  is an actual blow-up of  $C_5$  it is triangle-free so  $\alpha(G) \leq 2$ .

Observe first that  $e(G) = 2(k-1-a) + 2\binom{k-1}{2} + a^2 = k(k-1) + (a-1)^2 - 1$ , so

$$\frac{e(G) - 1}{|G| - 2} = \frac{k + 1}{2} - \frac{(k + 1)/2 - (a - 1)^2}{2k - 3} \leq \frac{k + 1}{2} - m(a).$$

Note that for  $k \geq 4$  we have  $\frac{(k+1)/2-(t-1)^2}{2k-3} \leq 1/2$  for any  $t \geq 1$ . This implies  $m(a) \leq 1/2$  and in particular since  $m_2(K_{k-1}) = k/2 \leq (k+1)/2 - m(a)$  it is enough to consider only proper subgraphs of size at least  $k$  or more when computing  $m_2(G)$  (or the desired bound holds). Let  $H$  be the subgraph of  $G$  maximising  $d_2(H)$  as in Lemma C.1. Since all vertices within a part are equivalent we know  $H$  must be a union of full parts. Above considerations tell us that  $H$  is not the whole graph and has at least  $k$  vertices.

Let us also observe that if  $\frac{a-1}{k-2} \geq 1/2$ , the second term is smaller (or equal) in the minimum for both  $m(a-1)$  and  $m(a)$  as well as being decreasing in  $t$  which

would make  $m(a - 1)$  larger than  $m(a)$ , a contradiction. Hence,  $a - 1 < (k - 2)/2$ , implying  $a \leq (k - 1)/2$  and in particular  $k - 1 - a \geq a$ . This implies that  $\delta(G) = \min(2(k - 1 - a), k - 1) \geq k - 1$ . If  $H$  has  $2k - 1 - t$  vertices with  $1 \leq t \leq k - 2$  we would have

$$\begin{aligned} d_2(H) &= \frac{e(H) - 1}{|H| - 2} \leq \frac{e(G) - t\delta(G) + \binom{t}{2} - 1}{2k - t - 3} \\ &\leq \frac{k(k - 1) + (a - 1)^2 - 1 - t(k - 1) + \binom{t}{2} - 1}{2k - t - 3} \\ &= \frac{k + 1}{2} + \frac{(a - 1)^2 - (k + 1)/2 - t(k - 3)/2 + \binom{t}{2}}{2k - t - 3} \\ &\leq \frac{k + 1}{2} - \frac{(k + 1)/2 - (a - 1)^2}{2k - 3} \\ &\leq \frac{k + 1}{2} - m(a), \end{aligned}$$

where in the first inequality we used the fact that  $H$  is obtained from  $G$  by removing a set of  $t$  vertices and hence, these  $t$  vertices are incident to at least  $\delta(G)t - \binom{t}{2}$  edges.

Finally if  $H$  has exactly  $k$  vertices, it is easy to see that it misses at least  $a$  edges (since it is union of parts of the blow-up and every part except the single vertex one has size at least  $a$ ) which implies  $\frac{e(H)-1}{k-2} \leq \frac{\binom{k}{2}-a-1}{k-2} = \frac{k+1}{2} - \frac{a}{k-2} \leq \frac{k+1}{2} - m(a)$ , completing the proof.  $\square$

The following lemma will be useful in determining when it makes sense to take vertex disjoint unions of graphs in computation of  $m_2(G)$ . For two graphs  $G$  and  $H$  we will denote by  $G \sqcup H$  the graph obtained by taking a vertex disjoint union of  $G$  and  $H$ .

**Lemma C.4** *Assume  $2e(H) > |H|$  and  $2e(G) > |G|$ . Then  $d_2(G \sqcup H) < \max(d_2(G), d_2(H))$ .*

**Proof** Assume otherwise, so  $d_2(G \sqcup H) \geq d_2(G), d_2(H)$ . We have  $d_2(G \sqcup H) = \frac{e(G)+e(H)-1}{|G|+|H|-2} \geq \frac{e(G)-1}{|G|-2} = d_2(G) \Leftrightarrow \frac{e(G)+e(H)-1}{e(G)-1} \geq \frac{|G|+|H|-2}{|G|-2} \Leftrightarrow \frac{e(H)}{e(G)-1} \geq \frac{|H|}{|G|-2} \Leftrightarrow \frac{e(H)}{|H|} \geq \frac{e(G)-1}{|G|-2}$ . Observe now that since  $|H| < 2e(H)$  we have  $\frac{e(H)}{|H|} < \frac{e(H)-1}{|H|-2} = d_2(H)$  so  $d_2(G \sqcup H) \geq d_2(G) \implies d_2(H) > d_2(G)$ . Repeating with  $H$  in place of  $G$  we obtain  $d_2(G) > d_2(H)$  which is a contradiction.  $\square$

The following result shows that bound of Theorem 1.7 is tight (up to constant factor in the constant term) for the first half of the modulae.

**Theorem C.5** *Let  $k = \lceil \frac{m}{r-1} \rceil$  and  $\ell = m - (k - 1)(r - 1)$ . Provided  $\ell \leq \frac{r-1}{2}$  and that  $k$  is sufficiently larger than  $r$  there exists  $c_r > 0$  such that  $M(m, r) \leq \frac{k+1}{2} - \frac{c_r}{\sqrt{k}}$ .*

**Proof** Let us first describe our graph  $G$ . We let  $G$  consist of  $\ell$  vertex disjoint copies of our graph  $G'$  with  $2k - 1$  vertices from Lemma C.3 and  $r - 1 - 2\ell \geq 0$  copies of

$K_{k-1}$ . The only information we will use about  $G'$  is that  $m_2(G') \leq \frac{k+1}{2} - \frac{1}{2\sqrt{k+1}}$  and that  $\alpha(G') \leq 2$ .

Observe first that  $\alpha(G) < r$  as any set of  $r$  vertices of  $G$  must have either 3 vertices in some copy of  $G'$  or 2 vertices in a copy of  $K_{k-1}$ , either way there is an edge and the set is not independent.

On the other hand observe that any graph  $H$  with  $\alpha(H) < r$  and more than  $3r$  vertices must satisfy  $2e(H) > |H|$ , since there can be at most  $(1 - 1/r)|H|^2/2$  non-edges by Turán's theorem so  $e(H) \geq |H|^2/(2r) - |H|/2$ .

Let us now take  $H$  to be a subgraph of  $G$  maximising  $d_2(H)$  and our goal is to show that, upon minor modification this  $H$  must live inside a single copy of  $G'$  or  $K_{k-1}$  which will let us complete the proof since we know both have small enough  $m_2$ .

We know that  $H$  must have at least  $k$  vertices or we obtain  $m_2(G) = d_2(H) \leq m_2(K_{k-1}) = k/2$ . This means that if  $H$  has at most  $3r$  vertices inside some copy of  $G'$  or  $K_{k-1}$ , we can remove them and this only changes the 2-density by  $O(\frac{1}{k}) \leq o(\frac{1}{\sqrt{k}})$ , where we are taking  $k$  sufficiently larger than  $r$ . Let us call the graph  $H_1$  obtained from  $H$  by deleting all vertices of  $H$  belonging to a copy of  $G'$  or  $K_{k-1}$  if there are at most  $3r$  of them inside that copy. Our new graph  $H_1$  by above observation has  $d_2(H_1) \geq d_2(H) - O(\frac{1}{k})$  and has the property that inside any copy of  $G'$  or  $K_{k-1}$  it has either none or more than  $3r$  vertices.

Applying the preceding lemma iteratively to restrictions of  $H_1$  to a single copy of  $G'$  or  $K_{k-1}$  we can find a subgraph  $H_2$  of either  $G'$  or  $K_{k-1}$  which satisfies  $d_2(H_2) > d_2(H_1) \geq d_2(H) - O(\frac{1}{k})$ . Since we know  $d_2(H_2) \leq \max(m_2(K_{k-1}), m_2(G'))$  and we know  $m_2(K_{k-1}) = k/2$  and  $m_2(G') \leq \frac{k+1}{2} - \frac{1}{2\sqrt{k+1}}$  so either way  $m_2(G) = d_2(H)$  is small enough.  $\square$

## References

1. Ajtai, M., Komlós, J., Szemerédi, E.: A note on Ramsey numbers. *J. Combin. Theory Ser. A* **29**(3), 354–360 (1980)
2. Albertson, M., Bollobás, B., Tucker, S.: The independence ratio and maximum degree of a graph. In: *Proceedings of the Seventh Southeastern Conference on Combinatorics, Graph Theory, and Computing* (Louisiana State University, Baton Rouge, LA, 1976), pp. 43–50 (1976). *Congressus Numerantium*, No. XVII
3. Alon, N., Spencer, J.H.: *The Probabilistic Method*, 4th edn. Wiley, New York (2016)
4. Alon, N., Sudakov, B.: On graphs with subgraphs having large independence numbers. *J. Graph Theory* **56**(2), 149–157 (2007)
5. Arora, S., Lovász, L., Newman, I., Rabani, Y., Rabinovich, Y., Vempala, S.: Local versus global properties of metric spaces. *SIAM J. Comput.* **41**(1), 250–271 (2012)
6. Balogh, J., Clemen, F.C., Lavrov, M., Lidický, B., Pfender, F.: Making  $K_{r+1}$ -free graphs  $r$ -partite. *Combin. Probab. Comput.* **30**(4), 609–618 (2021)
7. Bohman, T., Keevash, P.: The early evolution of the  $H$ -free process. *Invent. Math.* **181**(2), 291–336 (2010)
8. Bohman, T., Keevash, P.: Dynamic concentration of the triangle-free process. *Random Struct. Algorithms* **58**(2), 221–293 (2021)
9. Bollobás, B., Hind, H.R.: Graphs without large triangle free subgraphs. *Discrete Math.* **87**(2), 119–131 (1991)
10. Dudek, A., Rödl, V.: On the function of Erdős and Rogers, *Ramsey Theory. Progress in Mathematics*, vol. 285, pp. 63–76. Birkhauser/Springer, New York (2011)

11. Dudek, A., Retter, T., Rödl, V.: On generalized Ramsey numbers of Erdős and Rogers. *J. Combin. Theory Ser. B* **109**, 213–227 (2014)
12. Erdős, P.: Problems and Results in Combinatorial Analysis and Combinatorial Number Theory, Graph Theory, Combinatorics, and Applications, vol. 1 (Kalamazoo, MI, 1988), pp. 397–406. Wiley-Interscience Publications, Wiley, New York (1991)
13. Erdős, P., Rogers, C.A.: The construction of certain graphs. *Can. J. Math.* **14**, 702–707 (1962)
14. Erdős, P., Szekeres, G.: A combinatorial problem in geometry. *Compos. Math.* **2**, 463–470 (1935)
15. Erdős, P., Faudree, R.J., Rousseau, C.C., Schelp, R.H.: On cycle-complete graph Ramsey numbers. *J. Graph Theory* **2**(1), 53–64 (1978)
16. Erdős, P., Györi, E., Simonovits, M.: How many edges should be deleted to make a triangle-free graph bipartite? *Sets, Graphs and Numbers* (Budapest, 1991). *Colloquia Mathematica Societatis János Bolyai*, vol. 60, pp. 239–263. North-Holland, Amsterdam (1992)
17. Fiz Pontiveros, G., Griffiths, S., Morris, R.: The triangle-free process and the Ramsey number  $R(3, k)$ . *Mem. Am. Math. Soc.* **263**(1274), v+125 (2020)
18. Gromov, M.: Local and global in geometry. Arxiv preprint (2018). [www.ihes.fr/~gromov/wp-content/uploads/2018/08/1107.pdf](http://www.ihes.fr/~gromov/wp-content/uploads/2018/08/1107.pdf)
19. Hasse, H.: Darstellbarkeit von Zahlen durch quadratische Formen in einem beliebigen algebraischen Zahlkörper. *J. Reine Angew. Math.* **153**, 113–130 (1924)
20. Janzer, O., Gowers, W.T.: Improved bounds for the Erdős–Rogers function. *Adv. Combin.* **3** (2020)
21. Jones, K.F.: Independence in graphs with maximum degree four. *J. Combin. Theory Ser. B* **37**(3), 254–269 (1984)
22. Kamčev, N., Krivelevich, M., Morrison, N., Sudakov, B.: The König graph process. *Random Struct. Algorithms* **57**(4), 1272–1302 (2020)
23. Korándi, D., Roberts, A., Scott, A.: Exact stability for Turán’s theorem. *Adv. Combin.* **9** (2021)
24. Kostochka, A., Yancey, M.: Ore’s conjecture on color-critical graphs is almost true. *J. Combin. Theory Ser. B* **109**, 73–101 (2014)
25. Krivelevich, M.:  $K^s$ -free graphs without large  $K^r$ -free subgraphs. *Combin. Probab. Comput.* **3**(3), 349–354 (1994)
26. Krivelevich, M.: On the minimal number of edges in color-critical graphs. *Combinatorica* **17**(3), 401–426 (1997)
27. Krivelevich, M., Kwan, M., Loh, P.-S., Sudakov, B.: The random  $k$ -matching-free process. *Random Struct. Algorithms* **53**(4), 692–716 (2018)
28. Li, Y., Zang, W.: The independence number of graphs with a forbidden cycle and Ramsey numbers. *J. Combin. Optim.* **7**(4), 353–359 (2003)
29. Linial, N.: Local-global phenomena in graphs. *Combin. Probab. Comput.* **2**(4), 491–503 (1993)
30. Linial, N., Rabinovich, Y.: Local and global clique numbers. *J. Combin. Theory Ser. B* **61**(1), 5–15 (1994)
31. McKay, B.: Database of Ramsey graphs. <http://users.cecs.anu.edu.au/bdm/data/ramsey.html>
32. Osthus, D., Taraz, A.: Random maximal  $H$ -free graphs. *Random Struct. Algorithms* **18**(1), 61–82 (2001)
33. Roberts, A., Scott, A.: Stability results for graphs with a critical edge. *Eur. J. Combin.* **74**, 27–38 (2018)
34. Shearer, J.B.: A note on the independence number of triangle-free graphs. II. *J. Combin. Theory Ser. B* **53**(2), 300–307 (1991)
35. Staton, W.: Some Ramsey-type numbers and the independence ratio. *Trans. Am. Math. Soc.* **256**, 353–370 (1979)
36. Sudakov, B.: A note on odd cycle-complete graph Ramsey numbers. *Electron. J. Combin.* **9**(1), 4 (2002)
37. Sudakov, B.: Large  $K_r$ -free subgraphs in  $K_s$ -free graphs and some other Ramsey-type problems. *Random Struct. Algorithms* **26**(3), 253–265 (2005)
38. Sudakov, B.: A new lower bound for a Ramsey-type problem. *Combinatorica* **25**(4), 487–498 (2005)
39. Wolfowitz, G.:  $K_4$ -free graphs without large induced triangle-free subgraphs. *Combinatorica* **33**(5), 623–631 (2013)